

Uncomfortable truths

A thorough investigation of German scientists' actions under the Nazi regime reveals a more complex and ambiguous story than that implanted in the public mind at the end of the Second World War.

The inhumanity of some of the research carried out in Germany during the Third Reich is well known, and universally abhorred. The work of Josef Mengele, the young doctor who conducted deadly genetics experiments on inmates of Auschwitz, is perhaps the most notorious example of these appalling crimes.

For decades after the Second World War, the prevailing view of how scientists interacted with the Nazi regime was fixated on such cases of dramatic criminality. According to this view, science during the Nazi era was contaminated by a few, very rotten apples. This version of history also held that these rotten apples were engaged in 'pseudoscience' — low-quality research whose results were meaningless; that the Nazis held 'real' science in low esteem, so that the main body of scientists simply trod water for the duration; and that most of those who did work to further the aims of the regime did so under duress.

This conventional wisdom was broadly framed at the Nuremberg trials, which condemned the heinous crimes of high-ranking Nazis, but did not enquire into the behaviour of less notorious individuals, including rank-and-file scientists.

This account suited both the winners and losers of the war. By eliminating the worst offenders, justice was seen to be done. Experienced scientists and research managers were left alone to rebuild the science infrastructure of the destroyed country. The victorious Allies needed what was then West Germany to function as a strong but peaceful country, as a bastion against the communist threat to the East. And science was an integral part of that.

Second take

More recently, science historians have begun to question the rotten-apple/pseudoscience view, in parallel with a broader historical reassessment of this calamitous period in German history. The ambiguity and complexity of individual behaviour and motivation during the war have been reflected not just in academia, but also in plays such as *Copenhagen*, and a flurry of films including *Taking Sides*, *Sophie Scholl* and *Downfall* (*Der Untergang*).

Germany's main scientific institutions have been moved to reassess their own twentieth-century histories. For example, the university grant-giving agency, the DFG, has investigated its funding of research that supported Nazi policies. But it is the Max Planck Society (MPS), which administers 80 research institutes in Germany, that has taken the lead in exposing its own past to unflinching scrutiny.

In 1999, Hubert Markl, then MPS president, launched a six-year, €4-million (US\$5-million) programme, conducted by independent science historians, to systematically analyse the role of the society — then known as the Kaiser Wilhelm Society — and its scientists in supporting the Nazi regime's policies. The programme ended last month, and the results of its many projects confirm the superficiality of the accepted view.

The MPS has found that a large part of the most criminal research conducted was not 'pseudoscience' — in fact, it followed conventional scientific methods and was at the cutting edge of research at the time. It has also demonstrated that the Nazis held basic research in high esteem, increasing funding for it during the war years without

requiring scientists to join the Nazi Party. And it found that, far from being subjected to force, many scientists voluntarily oriented their work to fit the regime's policies — as a way of getting money and of exploiting the new resources that Nazi policies made available through, for example, the invasion of other countries. Most researchers, it turns out, seem to have regarded the regime not as a threat, but as an opportunity for their research ambitions.

Lessons for the future

It has taken more than 50 years for such a serious, dispassionate reanalysis to become possible, at least in Germany — for both psychological and practical reasons. First, a generation of academics is retiring, and their successors need a clear path, unburdened by the legacy of the Third Reich and the pressures to rebuild Germany after the war. Second, important Russian archives became available to Western historians only after the end of the cold war.

The programme's final conference, held last month in Berlin, made clear the productivity of the endeavour. A thick dossier of publications is also freely available on the website of the Max Planck Institute for the History of Science, which hosted the independent group in Berlin (www.mpiwg-berlin.mpg.de/KWG/engl.htm). The dossier portrays individuals who clearly overstepped the ethical line, such as plant geneticist Hans Stubbe, who collaborated with the SS to get hold of valuable Russian plant collections after the invasion of Russia.

It reveals even more about a large ethical grey area. Researchers at the Kaiser Wilhelm Institute for Metal Research in Stuttgart, for example, voluntarily came up with many projects to improve the performance of existing weapons. And it bears out accusations made against Adolf Butenandt, the Nobel-prizewinning biochemist who was president of the MPS during 1960–72 (see *Nature* **393**, 109–111; 1998). It seems that Butenandt must have known that Auschwitz blood samples were being handled at his Institute for Biochemistry in Berlin.

The exercise has also brought historians to the point where they can formulate new questions, moving on from the identification and condemnation of individuals to the more general issue of how the scientific community interacted with the regime, to better understand the history of that time and to learn from it.

In the 1930s and early 1940s, it seemed to those living under fascist flags that fascism was immortal. Until 1942, few Germans — or Italians, for that matter — imagined that the ruling regimes would be overthrown, or be replaced by a democratic system that would judge many of the actions they considered loyal, patriotic, or simply getting on with their job, as unacceptable support for a criminal regime.

The MPS is to be praised for its courage in opening itself up so completely to scrutiny, and for funding the investigation at a time when its own finances are being severely curtailed. The work has shed light on the behaviour of scientists as individuals and as groups. And it serves as a timely reminder of the need for strict ethical limits to be defined, and adhered to.

The action has been positive, and in its grim way liberating, for German scientists. "Those who are fixated only on the past will move blindly into the future," Markl told the programme's final conference, "as will those who are concerned only with the present." ■



Fighting back

US denies accusations
of design flaw in
nuclear warhead
p684

Career move

Change of job
provides refuge for
Japanese researcher
p685

Off the agenda

Plug pulled on study
to test health effects
of Agent Orange
p687



Arctic role

Fur flies as critics
attack rise in hunting
quota for polar bears
p688

Warning system steps up a gear for fresh Indonesian earthquake

Declan Butler

Warnings about the risk of a tsunami after the recent earthquake in Indonesia spread faster and more widely than they did for last December's calamitous event, officials in the region say.

The latest earthquake struck off the coast of Sumatra on 28 March, just days before an interim tsunami warning system for the Indian Ocean was due to come into force on 1 April. But warnings were still delivered much more rapidly than they had been in December.

In the event, the magnitude-8.7 earthquake didn't generate a large tsunami. That was just as well, because when it struck, only a handful of the 25 countries in the interim warning system had provided names and numbers for national points of contact. "The earthquake happened before our deadline for receiving contact points," says Keith Alverson, head of UNESCO's Global Ocean Observing System in Paris.

Under the interim system, the US Pacific Tsunami Warning Center (PTWC) in Hawaii

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

The 28 March disaster caused devastation on the island of Nias.

and the Japan Meteorological Agency will provide alerts on all seismic activity in the Indian Ocean region to round-the-clock contact points in the surrounding countries. The system, agreed in March (see *Nature* 434, 261; 2005), is intended as a stopgap until 2006, when agreement is due on the details of a full-blown warning system based on tide gauges and seafloor pressure monitors.

The PTWC itself responded much more

quickly and decisively than in December. Shortly after that quake it issued two bulletins — but only to members of the Pacific warning system. One, issued 15 minutes after the event, stated that a quake of magnitude 8.0 had occurred, with no risk of tsunamis to Pacific nations. The second, issued 45 minutes later, upgraded the quake to magnitude 8.5 and stated that there was "the possibility of a tsunami near the epicentre".

The centre's team then also attempted to contact colleagues in Indonesia and Thailand, both members of the Pacific system. It

was four-and-a-half hours before the PTWC sent a message to the Tsunami Bulletin Board — which goes by e-mail to international tsunami scientists and organizations — mentioning press reports of the disastrous tsunami.

The reaction was very different this time. Twenty minutes after the earthquake, the PTWC sent out a bulletin simultaneously to Pacific centres and to the bulletin board, warning that the event had "the potential to generate a widely destructive tsunami in the ocean or seas near the earthquake". It explicitly advised evacuating coasts within 1,000 kilometres of the epicentre.

The PTWC also alerted the US Department of State, which sent messages to US embassies in the Indian Ocean region. The embassies, in turn, informed local emergency management agencies. "The PTWC now pays particular attention to the Indian Ocean; last time they weren't looking at it," says Peter Pissierssens, head of ocean services at the Intergovernmental Oceanographic Commission in Paris.

But international organizations report that responses to these warnings were patchy. The authorities in some coastal areas did issue prompt alerts and evacuated coastal areas. And vibrations from the earthquake itself were enough to send many people running inland. "There has been some progress in getting a warning system, but not a huge amount," says Alverson.

Plotting the course of a quake

Researchers are hoping that data collected in the wake of the recent earthquakes in Indonesia will allow them to build better models of the relationship between different seismic events in the region.

Kerry Sieh, for example, a seismologist at the California Institute of Technology in Pasadena, is pulling together data from global positioning system receivers on and around Sumatra.

He hopes to study how the land moved before, during and after the 28 March event and the one last December that generated a tsunami.

Data coming in as *Nature* went to press showed some surprisingly large movements after the March earthquake, such as land at an airport on the nearby island of Simeulue, which rose 1.6 metres and shifted 2.3 metres towards the ruptured fault, says Sieh.

Sieh's data, taken continuously at various locations on the islands near the earthquake fault, should help to resolve a controversy over

whether changes in Earth's structure can be seen before an earthquake. Such land 'deformations' were not seen in the run-up to other well-monitored large earthquakes, such as Tokachi-Oki in northern Japan in 2003. Some scientists say that failure to see anything here would be another nail in the coffin for these efforts.

More positive results will probably come from mixing the geological observations with physical modelling of how earthquakes affect the surrounding region. In a 17 March paper in *Nature*, John McCloskey of the University of Ulster used models of how stress can move down faults to identify areas that could be at risk of rupture following the December earthquake (J. McCloskey, S. S. Nalbant and S. Steacy *Nature* 434, 291; 2005). McCloskey and Sieh now plan to work together to assess the risks of a rupture on faults south of the recent epicentres.

David Cyranoski, Tokyo

Pope praised for partial conciliation of science and religion

Quirin Schiermeier, Munich

Catholic researchers and bioethicists have responded to the death of Pope John Paul II with tributes to his efforts to achieve reconciliation between faith and science. And some are optimistic that his successor will keep on the same path.

The Polish Pope had a strong personal interest in science and worked to reduce hostility between the scientific community and the Roman Catholic Church. Nonetheless, his strict rejection of abortion, embryonic stem-cell research and contraception, including the distribution of condoms to help contain AIDS, drew him into conflict with some scientists.

AIDS activist groups around the world still condemn John Paul's refusal to endorse the use of condoms. "It should not be forgotten that millions have died in Africa as a result of this theological rigidity," the London-based *Independent* newspaper said in an editorial.

However, John Paul frequently discussed scientific matters with such luminaries as Stephen Hawking, who is one of the 80 members of the Pontifical Academy of Sciences. And the Vatican received regular scientific advice (see *Nature* 432, 669; 2004).

"Many of us have witnessed a special feeling between the Pope and scientists," says Giuseppe Tanzella-Nitti, a theologian and astrophysicist at Rome's Pontifical University of the Holy Cross.

In 1980 at Cologne Cathedral, Germany, John Paul declared that there was "no contradiction" between faith and science. He said on several occasions that the concepts of the Big Bang and darwinian evolution were more than mere hypotheses. In 1992, he officially rehabilitated Galileo Galilei, conceding that the Church was wrong to arrest him.

More recently, the Vatican has stopped opposing modern techniques such as organ transplantation and genetic modification of animals.

Ludger Honnefelder, a Catholic theologian and philosopher at the Institute of Science and Ethics in Bonn, Germany, claims that John Paul helped religion and science to coexist. He notes that the next Pope will have to deal with issues such as the implications of genetic modification in humans. "We expect well-balanced answers from the Church to new ethical challenges," Honnefelder says, "just as we expect science not to think of itself as an almighty system." ■

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Water bombs: US submarines such as the *Virginia* are armed with W76 thermonuclear warheads.

Nuclear chiefs scotch story on frailty of ageing warheads

Geoff Brumfiel, Washington

Senior weapons scientists emphatically denied a report this week that the most important US nuclear warhead has a design fault that could make it unreliable.

The New York Times reported on 3 April that the W76, a thermonuclear warhead launched from submarines, has a flaw in the design of its casing that could cause it to explode with much less force than expected — or not at all.

Leading scientists who designed the warhead say that the accusations are completely unfounded, however, and that the report's sources weren't heavily engaged in the design project. "There is nothing wrong with the W76," says Harold Agnew, who was director of the Los Alamos National Laboratory in New Mexico when the lab designed and tested the warhead 30 years ago.

Exact numbers are classified, but arms-control experts estimate that the W76 makes up almost one-third of the US stockpile of 10,000 warheads, as well as dominating the far smaller British nuclear arsenal. The warheads are compact and lethal: up to eight of them can sit in a single missile, each yielding an explosive force more than five times bigger than that of the bomb dropped on Hiroshima in 1945.

For about a year, some scientists have been expressing concern to government officials about the uranium case surrounding the warhead. The case is thin and light so that it can be carried on smaller, submarine-launched missiles. The critics claim that it might fail when the fission trigger of the bomb detonates, meaning the bomb's powerful fusion fuel would not ignite.

The bombs are "at best unreliable and probably much worse," Richard Morse, a retired plasma physicist from Los Alamos

told the *The New York Times*. Morse did not respond to *Nature's* requests for an interview.

Other scientists familiar with the weapon dispute Morse's assertions. "I think he's wrong," says Richard Garwin, a prominent former hydrogen-bomb designer. Agnew, who was present at several tests of the W76, says that it never failed to detonate. He adds that six to nine warheads are carefully analysed every year to make sure that ageing will not affect their performance.

"We have looked into this concern extensively, and our best technical judgement is that it is simply wrong," Linton Brooks told the Senate on 4 April. Brooks is the administrator of the National Nuclear Security Administration, which oversees the US nuclear stockpile.

But Brooks added that his agency is now planning a study to create a more robust type of warhead. This might eventually replace weapons such as the W76, which was designed for minimum size and weight. He says that the Reliable Replacement Warhead programme will design weapons that have wider performance margins and can be more easily maintained without testing.

Arms-control experts are sceptical of the project, which will cost \$9 million in 2006. "The existing stockpile is safe and reliable, and it is likely to be so for some time," says Daryl Kimball, executive director of the Arms Control Association in Washington, DC. "This programme seems unnecessary."

The replacement warhead programme would itself replace an 'advanced concepts' project, which has been criticized as a step toward development of new nuclear weapons (see *Nature* 428, 455; 2004). Opponents of new weapons note that doubts about the W76's reliability could boost political support for the replacement programme. ■

Job switch stymies Japan's abduction probe

David Cyranoski, Tokyo

A geneticist has landed a plum job at the Tokyo police department, just weeks after his work rekindled a row over the fate of Japanese citizens abducted by North Korea.

But critics are claiming that the transfer of Tomio Yoshii from Teikyo University to head the forensics unit at the Tokyo metropolitan police department is intended to shield him from enquiries about the accuracy of his DNA tests. In a fractious parliamentary exchange with foreign minister Nobutaka Machimura on 30 March, Nobuhiko Suto, a member of the opposition Democrat Party of Japan, suggested that the government had used its influence to plant Yoshii in his new position.

The Japanese government has claimed that Yoshii's DNA analysis proves beyond doubt that cremated remains provided by the North Korean government late last year were from someone other than Megumi Yokota — a Japanese woman kidnapped by North Korea in 1977. Japan has been demanding a full account of the fate of Yokota and several others who were allegedly kidnapped.

But in an interview with *Nature*, Yoshii has conceded that his results could have been the

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

DNA testing has been used to identify victims abducted by North Korea.

result of contamination (see *Nature* 433, 445; 2005). Japanese government officials have disputed the *Nature* article, claiming that Yoshii says he was misquoted. A documentary film-maker from Australia, a South Korean broadcasting company and other reporters have since sought, without success, to interview Yoshii.

Suto says he wants Yoshii to testify on the matter before parliament's foreign affairs committee. But in Yoshii's new position with the police force, he can only appear if his employer agrees — an arrangement that Suto says is being used as an obstacle. During

the 30 March debate, Suto told Machimura that it was "surprising" that "a civilian without real police training" should suddenly obtain a top position in the police department. "Isn't this just hiding a witness?" Suto asked.

Yoshii's results were taken as conclusive by the government despite contrary reports from the National Research Institute of Police Science in Chiba, which found that no DNA could be gleaned from the remains. "If we are going to take the word of a single academic at a private university over that of this huge research institute, shouldn't we just get rid of

the institute?" Suto asked Machimura.

Machimura called Suto's scepticism "an insult" and said that the ministry had taken the investigation seriously. "We were not just trying to pull out some predetermined conclusion," he protested. "I wish you would choose your words more carefully."

Suto still plans to call Yoshii to testify, and says he will get to the bottom of why the government set so much store by Yoshii's results. "If Japan keeps going in this direction," he warns, "it will undermine its scientific reputation."

Additional reporting by Junko Chikatani.

Y. SHIMBUN/NEWS.COM

Image problems jeopardize comet mission's impact

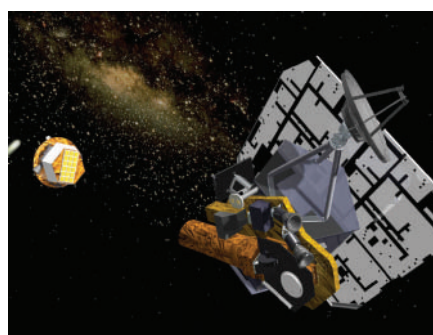
Rex Dalton, San Diego

Researchers are working against the clock to fix a blurry telescope whose malfunction could undermine a \$330-million US space mission to a distant comet.

The Deep Impact spacecraft is on course for comet Tempel 1, about 160 million kilometres from Earth. The car-sized craft has an 'impactor' that is due to split off and strike the comet on 4 July, while cameras on the main spacecraft capture the collision, enabling researchers to analyse the comet's make-up in unprecedented detail.

But Deep Impact's main instrument — a high-resolution telescope on the fly-by probe — is out of focus, admit project officials at the Jet Propulsion Laboratory (JPL) in Pasadena, California. This will affect the quality of the images captured from the experiment, the officials say.

The JPL project team has been working quietly for more than a month to try to fix the focus problem, which was discovered shortly after the probe was launched in Florida on 12 January. NASA, which is funding the project, acknowledged the



Out of focus: images of the Deep Impact craft's collision with comet Tempel 1 may be blurred.

problem on 25 March, after an enquiry from *Nature*.

JPL scientists have organized a special team to study possible remedies. But so far, efforts to correct the focus have failed, JPL officials say.

Project scientists now plan to employ computer imaging techniques that they hope will improve data from Deep Impact. Similar techniques have been successfully employed in the past on other

space telescopes, including Hubble, they say.

The problem telescope was built by Ball Aerospace & Technologies of Boulder, Colorado. Ball officials refused requests for an interview.

As the Deep Impact spacecraft approaches Tempel 1, the impactor will separate from the main craft. The impactor itself has an onboard telescope and camera, which will be used for navigation and for capturing images up to the point of collision.

The main spacecraft is due to fly past the comet at a distance of about 700 kilometres. It carries the problematic high-resolution telescope, as well as a back-up medium-resolution instrument, which is due to capture a wider view of the collision. The back-up instrument is functioning properly, officials say.

But even if pictures from the high-resolution telescope cannot be improved, the instruments are expected to provide "the most detailed pictures of a comet ever taken", says Michael A'Hearn, an astronomer from the University of Maryland at College Park who is chief scientist on the project.

JPL/NASA

Global health agency split over potential anti-terrorism duties

Erika Check, Washington

A revision of the rules that govern the World Health Organization (WHO) is being held up as countries disagree on how the agency should deal with suspected bioterror events.

The United States and its allies are pressing the WHO to take the lead in bioterror investigations. But other nations — especially poorer ones where the body has a vital role in public health — say this would involve the agency in questions of national security, and would compromise its political neutrality.

"This is a very big issue," says Barbara Rosenberg, an analyst at the Center for Arms Control and Non-Proliferation in Washington. "Countries are already reluctant to report disease and permit WHO access because of repercussions on trade and tourism. Throwing the security aspect in there will only make it worse."

Last November, a working group was unable to agree on how the WHO should tackle bioterrorist threats, and a meeting in February also failed. A last-ditch attempt to resolve the issue is scheduled for 12 and 13 May — just days before the negotiators are due to present the revised guidelines to the World Health Assembly, the agency's governing body.

A working draft of the international health regulations says that if the intentional release of a biological agent in a member country is suspected, that country shall "provide to WHO all relevant public health information, materials and samples". This would force cooperation with the WHO; currently, the agency must be invited to enter a country before it conducts an investigation there.

Rosenberg argues that this change will be seen as a threat, and make countries less inclined to cooperate. Others add that the agency does not have enough funds to take on such a role without compromising its public-health mission. However, given the lack of alternatives, many analysts believe the WHO must take some part in bioterror inquiries, and they predict that the negotiations will agree on a softer version of the current draft.

"For public-health and political purposes, there needs to be a process for carrying out that kind of investigation," says Michael Powers, an analyst at the Chemical and Biological Arms Control Institute in Washington. "If we continue to put it off we're likely to have to carry out one without procedures in place." ■



Flag waving: conservative David Horowitz is touring campuses to promote his academic bills of rights.

Professors bristle as states act to mould lecture content

Emma Marris, Washington

University faculty members in the United States are gearing up to oppose state bills that are being put forward by conservatives in the name of academic freedom.

Critics say that these 'Academic Bills of Rights', which are written to make sure that each side of an issue is presented in lectures at public universities, could in fact stifle academic freedom — and disrupt the teaching of science in contentious fields such as evolution and global warming.

"This would be a right-wing political takeover of the universities," says Tom Auxter, president of the United Faculty of Florida, the state's main academics' union.

Along with introducing protection from discrimination based on political or religious convictions, a bill being proposed in Florida calls on faculty members to refrain from introducing "controversial matter" unrelated to the course subject. It also requires them to present "serious scholarly viewpoints" other than their own.

Although the bill was written primarily with the humanities in mind, it would apply to all academic disciplines. On 22 March, Dennis Baxley (Republican, Ocala), who is backing the bill, said that it would make sure that alternatives to evolution are not shut out of universities.

"I do believe it has implications for the hard sciences," says Auxter. "It will waste a lot of time in the classroom because you will have to spend time covering a bunch of extraneous stuff — every crazy idea out there."

The American Association of University Professors (AAUP) is opposing similar bills

nationally, saying that faculty members should decide course content. "This effort is part of a larger pressure on higher education to politicize the agenda," says Ruth Flower, the AAUP's director of public policy.

David Horowitz, a marxist radical turned conservative activist, has written a template for the bills introduced in Florida and elsewhere. The Center for the Study of Popular Culture, a Los Angeles-based think-tank co-founded by Horowitz, has helped to establish campus-based groups to back the measure.

The campaign has gathered steam in recent weeks, with bills introduced in several states. Georgia passed a non-binding motion supporting the idea in March 2004, and Colorado dropped the bill only when major universities agreed to adopt its language at the administrative level. Other states, among them Maryland and Washington, have already rejected bills or put them on hold.

The AAUP also objects to a clause in Horowitz's draft of the bill that requires universities and professional societies to "maintain a posture of organizational neutrality with respect to the substantive disagreements that divide researchers on questions within, or outside, their fields of inquiry". Most states have dropped this clause, as they do not have jurisdiction over national societies.

According to the AAUP, Florida could be the first state to pass the bill. Baxley, a close ally of Governor Jeb Bush, says the outraged reception is evidence that academics are too inflexible. "I've been called an ass in the school newspaper at the University of Florida," he says, "and that demonstrates exactly what I am talking about." ■

US abandons health study on Agent Orange

Declan Butler

A programme to investigate the health and environmental damage caused by widespread use of the defoliant Agent Orange during the Vietnam War has been cancelled before it even began.

Scientists say that the collapse of the project is largely the result of cultural differences, a lack of communication, and a deep reservoir of suspicion between the Vietnamese and US governments.

The United States used Agent Orange to reduce forest cover during the Vietnam War. But since the war's end in 1975, Vietnam has suffered a high number of birth defects — estimated to be 2–3 times the expected number in some areas — which it blames on the defoliant.

The herbicides that made up Agent Orange were contaminated with dioxins, a highly toxic group of chemicals. But a lack of reliable epidemiological studies means that there is uncertainty over the suspected link between dioxins and birth defects. Such studies are difficult to do in part because a single test for dioxins costs US\$1,400.

The joint US–Vietnamese research project would have analysed dioxin levels in 300 mothers of babies with birth defects, along with 300 mothers of healthy children. The study was approved in May 2003 by the US National Institute of Environmental Health Sciences (NIEHS) based in Research Triangle Park, North Carolina. But the institute pulled the plug on the project last month because, after two years, the Vietnamese Ministry of Health had still not approved the research protocols needed to begin the work.



Vietnam blames many of its birth defects on Agent Orange.

David Carpenter, the study's principal investigator and an environmental health researcher at the University at Albany in New York, says that the project fell victim to politics with “two different cultures coming together and not communicating well”. This led to misunderstandings from the outset, he says. Before funds of \$1 million a year for the three-year project were freed up, the NIEHS provided \$300,000 for a pilot study to verify that dioxin levels would be detectable in Vietnamese women. Although done for

valid scientific reasons, this was not fully explained to Vietnamese officials, who viewed it as a snub, says Carpenter.

“The NIEHS was probably insisting on protocols to ensure a real, valid study; the implications of which the Vietnamese either didn't understand when they agreed, or else simply don't want,” says Jeanne Mager Stellman, a scientist at Columbia University in New York, whose research has provided maps of herbicide spraying in Vietnam (see *Nature* **422**, 649; 2003).

Anne Sassaman, a director at the NIEHS, defends the decision to cancel the project, saying that progress has been minimal despite repeated visits by NIEHS officials to Vietnam and the agency playing host to three Vietnamese delegations. A general agreement to conduct joint research between the countries (see *Nature* **416**, 252; 2002) is still in place, she adds, and the NIEHS “remains hopeful that other studies on Agent Orange can be conducted in the future; we would be very happy to support them”.

But researchers close to the programme say that the Agent Orange study was viewed by the NIEHS as a test case, and in the wake of its failure the agency is likely to be reluctant to entertain new proposals. “I'm not optimistic about what's next,” says Carpenter.

The study was expected to provide evidence for a class action suit on behalf of millions of Vietnamese plaintiffs against US manufacturers of Agent Orange. This case was dismissed by a US judge on 10 March on the grounds that use of the defoliant in Vietnam could not be considered a war crime. ■

Postdocs slam zealous attitude of NIH ethics office

Rex Dalton, San Diego

Is an honorary plaque costing little more than \$25 enough to cause a conflict of interest for a US National Institutes of Health (NIH) administrator?

The biomedical research agency's ethics office certainly thought so last year when it advised the National Postdoctoral Association (NPA) that it could spend only \$25 on an award plaque for Ruth Kirschstein, the former acting director of the NIH.

The ethics office, which deals with issues at the 27 NIH institutes and their 18,000 employees, faces a tough challenge. It is supposed to prevent lobbyists from corrupting the agency and to guard against insiders exploiting the system for personal gain. But critics say it is going too far.

That's certainly what Alyson Reed,

director of the Washington-based NPA, thought when her young organization sought to present Kirschstein with its inaugural Distinguished Service Award. “We had a hard time finding a plaque that cheap,” says Reed.

Holli Beckerman Jaffe, an attorney who directs the NIH ethics office, explains that only plaques of “little intrinsic value” are permitted. The \$25 was probably a rough estimate made by an office staffer, she says.

Another implication of the ethics rules became apparent last month when three top NIH officials — including Tony Fauci, director of the National Institute of Allergy and Infectious Diseases — were introduced by name, but not affiliation, at a dinner held by the lobby group Research!America.

The steady flow of such anecdotes is beginning to irk researchers both inside and

outside the NIH. “It's madness,” says John Hardy, chief of the neurogenetics lab at the National Institute on Aging. But he says the sometimes arbitrary rules are just something that researchers should learn to live with. “It is very hard for a large organization to have common sense,” he says. “You just accept it.”

But researchers may soon have to accept even more of it: last month the NIH published tighter conflict-of-interest rules (see *Nature* **434**, 3–4; 2005). These are already under attack from organizations such as the Federation of American Societies for Experimental Biology (FASEB).

“Many of the rules are overly restrictive,” says Paul Kincade, president of FASEB. He adds that their implementation will “limit the ability of NIH scientists to engage in critically important teaching and scholarly activity.” ■

US scraps study on cancer fallout from nuclear testing

Washington A study of the link between thyroid cancer and above-ground US nuclear tests has been cancelled by its funding agency.

The project, which was being led by Joseph Lyon, a preventive-medicine expert at the University of Utah in Salt Lake City, aimed to look for thyroid cancer and other ailments in 4,000 people who lived in Utah and Nevada in the 1960s. Above-ground experiments at the Nevada Test Site may have caused as many as 75,000 additional cases of thyroid cancer, according to an earlier study by the National Cancer Institute (see *Nature* **389**, 534; 1997).

Lyon's study, which was funded by the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, had already spent US\$8 million examining some 1,300 patients. He was informed of the cancellation on 21 March. Lyon says the study would take another \$6 million and three years to finish, but believes it was halted because of the government's reluctance to discuss the impacts of testing. The CDC says the decision was the result of a normal peer-review process.

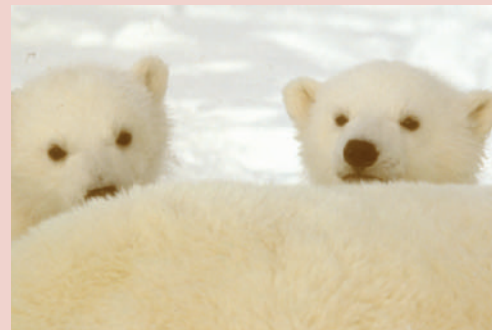
Zoologists get bear quota in their sights

Munich A hike in the annual quota of polar bears that can be killed by indigenous people and sports hunters in Canada could put the species at risk, say zoologists.

Canadian authorities decided in January to allow 518 bears to be killed this year, a 29% increase on 2004. The rise was granted at the request of indigenous hunters, who say that they have observed more bears this year than in recent years.

Some researchers say that the quotas are not based on the best available scientific data and were set without proper consultation with experts in other countries.

Officials counter this, saying that scientific studies were considered, and adding that traditional Inuit knowledge about population size



deserves more trust than it has had in the past. But Øystein Wiig, a mammalogist at the University of Oslo, told *Nature* that: "The harvest could be much higher than the populations in some areas can actually take."

Female physicists get 'expected' share of jobs

Washington Family obligations and hostile working environments are not to blame for the small number of female researchers in physics departments, claims an analysis by the American Institute of Physics.

The study, which looked at physics from high-school to university level, says that the number of women entering the field is small, but that those who sign up are

no more likely to leave than their male colleagues. Some observers blame the low number of women — who make up just 10% of physics staff at US degree-granting departments — on factors such as a lack of family-friendly policies. "The most surprising thing is that women are on the faculty at about the rates we would expect" given the starting numbers, says Rachel Ivie of the institute's statistical research centre, based in College Park, Maryland, which published the report in February.

But she adds that less than one-quarter of physics bachelors' degrees are awarded to women. "It is still primarily considered a man's occupation," says Ivie.

Former president finds physics niche in Russia

Moscow Askar Akayev may have been run out of his job and his country during popular uprisings last month, but the former president of Kyrgyzstan is being provided with a soft landing — the Russian science community.

A former physicist, with 80 publications to his name, Akayev has been offered a position at the Institute of Electrophysics in the Urals branch of the Russian Academy of Sciences in Ekaterinburg. "For

politics he is too intelligent and soft," says Gennady Mesyats, director of the institute. "But in the scientific community Akayev is an authority."

Akayev left Kyrgyzstan on 24 March after his offices in the capital Bishkek were stormed by protestors demonstrating against corruption and the disputed elections in February. Unlike most scientists in Russia, he is unlikely to be troubled by the task of living on a meagre salary: he has considerable property and business interests in central Asia.

Defence doyens push Bush to invest in green energy

Washington A group containing some high-profile conservatives has called for the United States to wean itself off foreign oil by investing in energy-efficiency measures and alternative fuels such as biomass.

The group wrote to President George W. Bush on 24 March to ask for \$1 billion of investment over the next five years. Green groups have long argued for increased spending on technology such as alcohol-fuelled and electric cars on environmental grounds, but this group argues that US reliance on oil threatens the country's commercial and security interests.

Authors include William Crowe, chairman

of the Joint Chiefs of Staff between 1985 and 1989, as well as John Dalton, former Secretary of the Navy, and national security experts from several US think-tanks.

Lockheed locks horns over Los Alamos lab

San Diego Technology giant Lockheed Martin has entered a race to manage Los Alamos National Laboratory in New Mexico.

There are only two bidders for the lab, which has an annual budget of \$2.2 billion and houses over 10,000 employees. The other is its current manager, the University of California, which has been humbled in recent years by security missteps and management strife (see *Nature* **433**, 447; 2005). The university's contract expires on 30 September this year.

Lockheed Martin, which is based in Bethesda, Maryland, says that changes to the lab's employee pension plans and management structure prompted it to enter the running late last month after declining to do so in 2004.

The move is likely to generate mixed feelings among lab personnel. Many want to see a competitive process, but staff also value the academic atmosphere generated by having a university, rather than an industrial company, manage the facility.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Out: Kyrgyzstan ditched its president in March.

Melioidosis? Never heard of it...

Deadly tropical infections that kill within 48 hours don't usually go unnoticed. But one killer has been largely ignored for decades. Now, thanks to worries about bioterror, it is being taken more seriously. Peter Aldhous reports.

Mention melioidosis in most circles — even those with a passing interest in tropical medicine — and you'll be met with blank stares. The infection is often misdiagnosed because the bacterium that causes it, the soil-dwelling *Burkholderia pseudomallei*, triggers multiple symptoms that mimic those of other diseases. In parts of Asia where *B. pseudomallei* is endemic, this serial killer often commits its crimes without even being identified as a suspect.

Not so in Ubon Ratchathani, a bustling provincial capital in northeast Thailand. For Wipada Chaowagul, a specialist in internal medicine at the city's Sappasitprasong Hospital, melioidosis is public enemy number one. Each year, the hospital admits about 200 people who test positive for *B. pseudomallei*. Up to half of them die.

Chaowagul's patients are mostly rice farmers. When the rains come each year, between May and October, *B. pseudomallei* threatens anyone paddling in the flooded paddy fields that surround Ubon — especially those already weakened by other conditions such as diabetes. Some develop internal abscesses or inflamed joints; others have difficulty breathing. Many are overwhelmed by the infection, and die from septic shock within 48 hours.

Chaowagul wants to run clinical trials of

new antibiotics, to see if they can reduce this toll. But for Thailand's overstretched health-care system, such drugs are prohibitively expensive. No pharmaceutical company has volunteered to donate its products, so Chaowagul's plans remain stalled. "If we use our own money, we have a problem," she says.

In from the cold

This tale is echoed for 'orphan' diseases across the developing world — unless pathogens afflict rich Westerners, they tend to attract little research money. But *B. pseudomallei* may soon lose its orphan status, thanks to fears that it might be used as a biological weapon. Through its richly funded biodefence initiative, the US National Institute of Allergy and Infectious Diseases (NIAID) is now encouraging microbiologists to begin working on the bacterium. "We're looking at building a research base," says Michael Schaefer, an official at the NIAID's headquarters in Bethesda, Maryland.

There are many puzzles to solve. *Burkholderia pseudomallei* is a resilient organism, able to hunker down in the soil or inside the cells of its human victims for years on end, only emerging when conditions favour its growth. One US Vietnam veteran, probably infected after breathing in aerosols of *B. pseudomallei* whipped up by helicopters,

first became sick 26 years later. Vanaporn Wuthiekanun, who works on *B. pseudomallei* in the Wellcome Trust unit at Mahidol University in Bangkok, has cultured the bacterium from a sample kept in distilled water for a decade. "It's very tough," she observes.

For the most part, *B. pseudomallei* is thought to get its nutrition from rotting organic matter, and when the opportunity arises, by parasitizing soil-dwelling amoebae. Its ability to infect human cells may simply be an unhappy consequence of the mechanisms that allow it to do the latter. But these mechanisms are poorly understood, as are the ecological factors that influence *B. pseudomallei*'s distribution across the tropics. One mystery is why it is absent in central Thailand — where it is replaced by its cousin, the harmless *B. thailandensis*.

"It's probably something to do with the soil, but we haven't worked it out yet," says Nick Day, who heads the Wellcome Trust's Bangkok unit. As if the scientific challenges weren't enough, researchers out in the field also face obstacles imposed by southeast Asia's history of conflict. "We're keen to do soil surveys in Cambodia, but we're afraid of the landmines," says Wirongrong Chierakul, who works in Day's unit.

From the limited information available, it is clear that *B. pseudomallei* is present in the

soil in parts of Asia that record few cases of melioidosis. Almost certainly, this is because of inadequate medical diagnosis. "The distribution of cases tends to follow the distribution of decent microbiology labs, of which there are remarkably few," says Day.

In places where melioidosis has been recognized as an important public-health issue, the priority is finding more effective treatments. Having evolved to compete in the soil with organisms that secrete antibacterial compounds, *B. pseudomallei* is resistant to many drugs. Until the mid-1980s, melioidosis was treated in Thailand with a cocktail of four conventional antibiotics. Only about one in five patients pulled through.

That was when Chaowagul teamed up with Nick White, who now directs the Wellcome Trust's southeast Asia programme, but was then heading up its Bangkok unit. The trust, Britain's largest biomedical research charity, has a long-standing interest in tropical diseases, and so agreed to launch a clinical trial to test a newer antibiotic, called ceftazidime. It halved the death rate¹. The drug's high cost posed a problem, but has since been lowered. "If you have the results of a trial, you can create a political snowball," says Day. "Very often the prices will come down."

Drug shortage

But Chaowagul still loses more than 40% of her patients. So she and White next turned to a class of antibiotics called the carbapenems, which in lab tests seemed to be effective against *B. pseudomallei*². But carbapenems are even more expensive than ceftazidime, and manufacturers were reluctant to donate them for trials. It wasn't until 1994 that White won a supply of a drug called imipenem from the US-based giant Merck, after writing a personal letter to Roy Vagelos, the company's chief executive.

The initial trial yielded encouraging but inconclusive results. Although there were fewer treatment failures than for ceftazidime, the sample size was too small to confirm whether the death rate was reduced³. The researchers wanted to press ahead with a larger study. But by then Vagelos had left Merck, and White's requests for further donations of imipenem fell on deaf ears. With the drug costing about US\$100 per patient, per day, the trial was doomed. "We just couldn't afford it," says White.

Allen Cheng, a clinician at the Menzies School of Health Research in Darwin, Australia, recounts similar difficulties. He is running a trial in Ubon to see whether Thai melioidosis patients are helped by granulocyte colony-stimulating factor (G-CSF), a signalling molecule that can boost the production of white blood cells. "This is virtually a self-funded project," says Cheng. The firms he approached showed little interest.

Melioidosis is also a local problem in northern Australia. There, G-CSF is used



Vanaporn Wuthiekanun visits a Thai rice field — the favourite haunt of a bacterial killer.

routinely, and the death rate has been reduced to less than 20%. But it's unclear whether similar results could be achieved in Thailand. Australian melioidosis patients are rushed into intensive care. But in Ubon, those infected with *B. pseudomallei* must take their chances on overcrowded general wards. At least the melioidosis patients get beds near to the nursing stations. Those with less serious conditions are jammed into corridors, on balconies and outside the elevators.

Against this gloomy background, the NIAID's biodefence initiative has provided a beacon of hope. *Burkholderia pseudomallei*

has not yet been used as a biological weapon, but its close relative, *B. mallei*, was used as a biological agent in the First World War. It causes glanders, a disease that kills horses and, more rarely, people. The bacterium was spread by German troops in an attempt to disable the Russian army's horses and mules.

Potent weapon

Given the potential for *B. pseudomallei* and *B. mallei* to be deployed by bioterrorists, the NIAID is keen to promote research on both, and last August held a meeting in Bethesda to kick-start interest. Those who have for years ploughed a lonely furrow in melioidosis research can scarcely believe the shift in gear. "Two or three years ago, no one else was working on this disease in North America," says Donald Woods, a microbiologist at the University of Calgary in Alberta, Canada. "Now there are ten labs."

The time is right for an influx of money and personnel. Last year saw the publication of the complete genome sequence of *B. pseudomallei*⁴, a product of the Wellcome Trust's enthusiasm for tropical medicine and genomics. It appeared next to another paper⁵ describing the genome of *B. mallei* — a project backed by the NIAID. The two sequences should provide clues for researchers trying to understand the organisms' virulence, and suggest targets for drug development.

The genome projects have also revealed that *B. mallei* evolved directly from *B. pseudomallei*, losing parts of its genome along the way⁶. For clinicians, this close relationship is encouraging. With funding from the US military and the Canadian government, Woods has developed a vaccine against glanders. In December, he began tests to see whether it can prevent horses from becoming infected with *B. mallei*. If the results are positive, Woods has high hopes that the vaccine will also protect people against melioidosis.

Maybe so, but those who have struggled to launch clinical studies in Thailand are unsure about the extent to which the biodefence bonanza will yield tangible benefits for them and their patients. For Cheng, one question overrides all others: "How do we stop people dying?"

NIAID officials say that the initial focus will be on basic research. But their goal is to use this knowledge to develop improved diagnostic tools, treatments and vaccines. And Schaefer says that the agency will consider funding clinical trials, if a strong case can be made. "The scope of the biodefence effort is very broad," he says.

Peter Aldhouse is Nature's chief news & features editor.

1. White, N. J. *et al.* *Lancet* **ii**, 697–701 (1989).
2. Dance, D. A., Wuthiekanun, V., Chaowagul, W. & White, N. J. *J. Antimicrob. Chemother.* **24**, 295–309 (1989).
3. Simpson, A. J. H. *et al.* *Clin. Infect. Dis.* **29**, 381–387 (1999).
4. Holden, M. T. G. *et al.* *Proc. Natl Acad. Sci. USA* **101**, 14240–14245 (2004).
5. Nierman, W. C. *et al.* *Proc. Natl Acad. Sci. USA* **101**, 14246–14251 (2004).
6. Godoy, D. *et al.* *J. Clin. Microbiol.* **41**, 2068–2079 (2003).

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Tough opponent: *Burkholderia pseudomallei* is a particularly resilient soil bacterium.

After the gold rush

California's voters have authorized the spending of \$3 billion over the next decade on stem-cell research.

But will this bonanza bring threats as well as opportunities? Peter Aldhous weighs the hopes and fears.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Bob Klein is a man in a hurry. This energetic 59-year-old entrepreneur rose to public prominence last year, spearheading the campaign that convinced California's voters to back an audacious plan to create a \$3-billion public fund to advance research on human stem cells. Interviewed by *Nature* in late February, he was determined to press ahead as soon as possible: "My goal would be to approve the first grants in May."

But today Klein and his supporters are wrestling with the practical realities of building their multibillion-dollar research agency, called the California Institute for Regenerative Medicine (CIRM), from scratch, while fending off political and legal attacks from both long-standing opponents and former allies. It's already clear that Klein's May deadline will be missed by several months.

Stem-cell biologists are looking forward to the injection of funds, whenever it finally

arrives. But some question whether the CIRM can spend the huge sums at its disposal without wasting money on second-rate research. Some also fear a public backlash against stem-cell research, as claims about 'cures' made during the ballot campaign may have left people with unrealistic expectations. Others worry that the initiative has set a dangerous precedent by moving the centre of gravity of an important area of biomedical research away from the federal government, and subjecting it to the shifting winds of state electoral politics. Indeed, with several other states following California's lead and providing budgets for stem-cell research (see map, opposite), an unprecedented shift in the balance of power is under way. And biologists are divided on its likely consequences.

The CIRM's legal and political difficulties centre on its governance, which is unusual for a state agency. The stem-cell ballot initia-

tive, known as Proposition 71, placed the agency under the control of a 29-person citizens' committee with little input from California's legislature. The committee, chaired by Klein, includes representatives of the biotech industry and academic institutions likely to receive CIRM funding.

Under attack

Lawsuits have been filed denouncing the CIRM as unconstitutional. These have been denied a hearing in the California Supreme Court, but the battle will resume in lower courts, and could yet interfere with the institute's plans to begin distributing money. Meanwhile, state senators Deborah Ortiz, a Democrat, and George Runner, a Republican, have said that they will press for a new ballot initiative to "reform" the stem-cell programme. Ortiz backed Proposition 71, but has since criticized the initiative's procedures for accountability

Washington A proposed law, which has passed one chamber of the state legislature and will be up before the other by mid-April, would ban cloning for reproduction but support it for research. Democratic governor Christine Gregoire — voted into office last autumn by the skin of her teeth — supported embryonic stem-cell research during her campaign.

Wisconsin It was here that biologist James Thomson did the pioneering work that made the field of human embryonic stem-cell research possible. Last November, Governor Jim Doyle proposed a massive spending boost for biotechnology, with a \$750-million package for work including stem-cell research. The proposal includes \$375 million for a Wisconsin Institute for Discovery at Thomson's home base, the University of Wisconsin at Madison.

Illinois State officials hope to create a \$1-billion Illinois Regenerative Medicine Institute this year, which would do stem-cell work and be financed by a tax on elective cosmetic surgery. The measure passed an initial hurdle in the state legislature on 17 March, but still has a long way to go.

New York Bills in favour of embryonic stem-cell research have died in the Republican-controlled Senate over the past few years. The Assembly has now proposed spending \$100 million in 2006 on the work. The measure faces an uphill battle which should play out by June.

Massachusetts With some of the most respected stem-cell researchers in the world, and a multimillion-dollar private stem-cell research foundation already set up, the Bay State looks set to pass a law explicitly supporting stem-cell research. In late March, the House passed a bill authorizing therapeutic cloning for medical research by a two-thirds majority — enough to override a veto expected from the state's Republican governor.

Connecticut The governor has proposed \$20 million over two years for stem-cell research. Scientists at Yale University and the University of Connecticut are lobbying for even more — up to \$100 million over ten years.

New Jersey Last year, New Jersey beat California to the punch to create the first state-funded, dedicated stem-cell research centre, the Stem Cell Institute of New Jersey. Building should begin this August. The state has already devoted \$11.5 million to the institute, and the governor is lobbying for an additional \$380 million for the next seven years.

Maryland Lawmakers are considering a \$23-million spending bill for embryonic stem-cell research. As *Nature* went to press, the bill had passed the House but not the Senate.

Virginia The state's legislature has established a research fund in the name of actor and stem-cell research advocate Christopher Reeve. But it will not fund work on human embryonic stem cells.

California Proposition 71 has devoted \$3 billion over the next ten years to stem-cell research.



- Legislation prohibits cloning for research purposes
- Legislation on cloned embryos for research unclear or absent
- Legislation supports therapeutic stem-cell research, including on cloned embryos

Source: National Conference of State Legislatures, March 2005

and avoiding conflicts of interest.

Assuaging these concerns will test Klein's powers of persuasion. But there's no doubting his personal commitment to the cause. Having made his fortune providing quality low-cost housing in the state, Klein ploughed \$3 million of his own money into the Proposition 71 campaign. And he makes no secret of his hope that stem-cell research will help his 14-year-old son, who has juvenile diabetes. Klein recalls a 2002 conversation with his son: "He said, 'Dad, don't worry about me. Everyone is dying. I'm just dying a little faster.' For a father, that's not something you can live with."

Klein's heart-on-sleeve demeanour and boundless energy were just what was needed in winning support for Proposition 71. But setting up a new research agency demands the skills of a consummate scientist-administrator, which he is not. Headhunters are still drawing up a shortlist of candidates for the key position of president of the CIRM. In the

interim, the post has been taken by Zach Hall, whose previous roles include directing the National Institute of Neurological Disorders and Stroke in Bethesda, Maryland.

Hall has already begun to temper Klein's enthusiasm with a dose of caution, noting that the practical challenges of setting up a review process rule out the initial goal of approving some grants in May. "We still don't have our working groups appointed," says Hall. "In the fall we might be able to approve our first grants."

The CIRM's main goal is to support research that either can't be backed by the federal government, or is unlikely to receive "timely or sufficient" funding. Although the US Congress has recently considered the question of loosening federal rules¹, the regulations currently limit federal dollars to work on the 78 human embryonic stem-cell lines created before August 2001. Many of the lines that seem most promising for

research were isolated after this date. Researchers also say they need new cell lines that haven't been grown on a layer of mouse cells, as this causes biochemical contamination that will rule out their eventual use in human therapies².

So the CIRM's top priorities will include the creation of new embryonic stem-cell lines, mostly from embryos left over from *in vitro* fertilization procedures, but also from embryos created by cloning. In addition to investigating the feasibility of 'therapeutic cloning', in which a patient's own cells might be used to derive grafts to replace damaged tissues, cloning can be used to create embryonic cell lines from patients with genetic diseases — which could prove invaluable for understanding these conditions and testing potential treatments^{3,4}.

Biologists also want to investigate the basics, such as how stem cells arise in the developing embryo, and which signals coax

them into developing into different tissue types. "I think everybody agrees that more fundamental work needs to be done," says Evan Snyder, a stem-cell biologist at the Burnham Institute in La Jolla. "This is an initiative to move the field forward in an unencumbered way."

Researchers will have to be careful to ensure that work involving non-approved cell lines is strictly separated from projects involving federal grants. In the long run, the CIRM may provide funding to build new labs to help this separation. But Klein wants to avoid delay, by leasing space lying empty in the San Francisco Bay Area and other cities. "There are lots of opportunities," he says.

Maybe so, but experts in the field point out that relatively few researchers currently have expertise in working with human embryonic stem cells. And this, some biologists argue, makes it unwise to pour hundreds of millions of dollars into this area any time soon. "A lot of money could be wasted on what is not the best work," warns Sean Morrison, a stem-cell biologist at the University of Michigan in Ann Arbor.

Institute officials say that maintaining quality will be their priority, adding that their efforts won't be entirely focused on embryonic stem cells. The CIRM will support work on adult stem cells, and fund research into other questions that must be addressed to bring stem-cell therapies to the clinic. The agency might also choose to back specific therapeutic strategies that have been neglected by the federal National Institutes of Health (NIH), Klein argues.

Firm foundations

Klein adds that researchers have difficulty winning grants from the NIH in a young field such as stem-cell research, as it's difficult for them to gather enough preliminary data to make a case for serious study. "Some of the brightest minds cannot get access to funding," says Klein. So part of the CIRM's strategy will be to provide 'seed' money to allow researchers new to the field to gather sufficient results for a convincing full grant application. And the agency's first awards are expected to be for academic institutions across California to hire graduate students and postdocs. "What we're calling for here is a vast expansion of the intellectual and scientific workforce," says Hall.

Biologists welcome this emphasis on bringing young researchers into the field. But some are anxious about the CIRM's wider effects on the US funding landscape. With other states pitching in, the momentum in stem-cell research is shifting away from the federal government — and opinion is divided on whether that's a good or a bad thing.

"This represents a paradigm shift in how biomedical research gets funded in this country," says Snyder. "It could turn out to be a really wonderful opportunity." But other

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Too much, too soon? Some scientists are worried that the election campaign for Proposition 71 oversold the state of stem-cell research.

biologists see potential threats. For one thing, state activity is likely to provoke opponents of embryonic stem-cell research in the federal government into further action. 'Pro-life' conservatives in the US Congress will almost certainly resume legislative attempts to ban all human cloning procedures, whether for reproduction or research.

But the deeper concern is that the Bush administration and its congressional allies will let the states take the lead, and scale back funding for the NIH. The agency's budget is already under pressure, given Bush's programme of tax cuts and the cost of the Iraq war, and the biomedical research community knows that it can't expect many favours. "I am worried that the willingness of states to take up the slack is going to absolve the federal government of its responsibilities in this area," says George Daley, a stem-cell researcher at Harvard Medical School in Boston.

Many biologists are concerned about the implications of the NIH having a reduced role in stem-cell research. They respect the federal agency's procedures for backing scientific excellence and adhering to rigorous ethical standards. And although Hall says that the CIRM plans to import many NIH procedures, the institute and the other state funding mechanisms are still unknown enti-

ties. "The less the NIH is involved in something, the less assurance we have that good ethical guidelines are followed," observes Jon Retzlaff, director of legislative relations with the Federation of American Societies for Experimental Biology.

Hope or hype?

Another worry about putting stem-cell research in state hands is that it may be more vulnerable to the ebb and flow of public opinion. And this, some researchers fear, may be set for a backlash. The use of the slogans "cures for California" and "save lives with stem cells" during the Proposition 71 campaign may have set public expectations too high, they say. Disappointment could result if treatments fail to emerge rapidly, or if patients suffer adverse events in clinical trials — as has happened in the field of gene therapy⁵.

"I think there was a concern about promising too much, even before Proposition 71," says Gordon Keller, a stem-cell biologist at the Mount Sinai School of Medicine in New York. "My fear is that there are people who will try to move forward quicker than they should, and that could backfire very badly."

The media are already alert to this possibility. A 23 February article in *The New York Times*, for instance, questioned whether a planned clinical trial to use human embryonic stem cells to treat spinal injury is going "too far, too soon into uncharted territory". The trial, which could begin next year, is based on the work of Hans Keirstead of the University of California, Irvine, who has shown that embryonic stem cells can be made to differentiate into cells that help to maintain and protect neurons⁶.

Keirstead took a prominent role in the Proposition 71 campaign, discussing unpublished experiments in which spinal-injured rats were made to walk again. "It has never been my intention to contribute to the hype," says Keirstead, who stresses that he doesn't expect to see similarly dramatic results in human patients. But such subtleties were lost in the frenzy of the campaign. Some Californian stem-cell researchers say that they are now planning outreach activities to help manage public expectations.

Indeed, some researchers who are gearing up to receive CIRM funding feel a heavy burden of responsibility. "A lot of people do feel accountable here in California," says Fred Gage, a stem-cell researcher at the Salk Institute for Biological Studies in La Jolla. "We want to make sure that it's done right, because we're under the microscope."

Peter Aldous is Nature's chief news & features editor.

1. Check, E. *Nature* **434**, 548 (2005).
2. Martin, M. J., Muotri, A. Gage, F. & Varki, A. *Nature Med.* **11**, 228–232 (2005).
3. *Nature* **414**, 567 (2001).
4. Daley, G. Q. *N. Engl. J. Med.* **351**, 627–628 (2004).
5. Check, E. *Nature* **420**, 116–118 (2002).
6. Nistor, G. I., Totoiu, M. O., Haque, N., Carpenter, M. K. & Keirstead, H. S. *Glia* **49**, 385–396 (2005).

Arrogance imperils plans for change at Harvard

An abrasive president may find it counterproductive to alienate so much of the faculty.

Sir—I was disappointed by your recent Editorial (“Why Harvard needs Summers” *Nature* **434**, 1; 2005), published in the aftermath of Harvard president Larry Summers’ suggestion that women fail to advance in science because they are innately less able than men. His comments have done incalculable harm, contributing greatly to the hostile environment that causes women to leave science.

Summers’ views are opposed by an avalanche of data showing that women are as capable as men in science, but often cannot succeed by merit alone because of prejudice. One study (C. Wennerås and A. Wold, *Nature* **387**, 341–343; 1997) found that women applying for a research grant needed to be 2.5 times more productive than men in order to be considered equally

competent; for many more, see *Why so Slow?* by Virginia Valian (MIT Press, Cambridge, MA, 1999). No wonder women are not succeeding! Summers’ views amount to blaming the victim.

As Stephen Jay Gould’s wonderful book *The Mismeasure of Man* (Norton, New York, 1996) shows, theories about the supposed innate inferiority of women and minorities invariably derive from social prejudice. Many well-meaning people have these biases and are unaware of them. We all need to be more aware of our social biases, and we all need to speak out and confront sexism and discrimination whenever we encounter them.

For this reason, I have been disappointed by the failure of our, largely male, scientific leadership to speak out about the inaccuracy

of Summers’ comments. “*Qui tacet consentire videtur*: he who keeps silent is assumed to consent” — and the silence has been deafening. It is difficult for women scientists: if they speak out, they are viewed as asking for undeserved benefits, whereas if they keep silent, progress cannot be made. That’s why I think the MIT professor who brought Summers’ comments to public attention, Nancy Hopkins, is a hero.

At this point, Summers’ arrogant and unscholarly approach has so deeply antagonized the Harvard faculty that there is little chance he can achieve the positive changes enumerated in your Editorial.

Ben A. Barres

Department of Neurobiology, Stanford University School of Medicine, 299 Campus Drive, Stanford, California 94305-5125, USA

DNA barcoding is no substitute for taxonomy

Sir—DNA barcoding, formerly a way of identifying DNA within foodstuffs, is now being proposed as a way to catalogue life.

At the First International Barcoding Conference, held at the Natural History Museum in London in February 2005, heads of research institutes discussed plans to use museums, herbaria and other biodiversity institutes as national centres for DNA barcoding. Several small grants promoting it as an economical way of cataloguing life have been awarded, with a view to seeding bids for larger consortium grants.

Claims for its benefits are extravagant. The Consortium for the Barcode of Life has stated: “DNA barcoding will make a huge difference to our knowledge and understanding of the natural world.” But that’s a slogan. What of the science behind these undertakings?

The purpose of barcoding is to find a unique piece of DNA (cytochrome *c* oxidase subunit 1, for example) for every described species, so future taxonomists can run large biotic surveys without the need to learn or use morphological keys. Barcoding is at best a technology that may be able to spot DNA diversity within physically indistinct species. But even at this level it remains a genetic key to identify known species, rather than replacing traditional taxonomic practice.

However, this quick, cheap technology is in competition with taxonomy for funding. What cash-strapped student will want to enter a field such as taxonomy that

takes years to master and offers little or no job prospects? A budding barcoder — with no interest in biology, let alone taxonomy — can be trained in a fraction of that time, quickly disseminate their ‘research’ globally and look forward to a well-funded career.

DNA barcoding may seem progressive to those who use the word ‘dusty’ whenever the subject of taxonomy arises. But the work of taxonomists provides knowledge of the organism, not a few possibly unique nucleotides. In any case, every barcode must be linked with a known, described specimen stored somewhere.

Given its high-profile launch, barcoding will almost certainly result in a plethora of newly ‘flagged’ DNA species that will never be formally described. One estimate is that it will take some 250 years for taxonomy to catch up with barcoding. True to form, barcoding has supplied an answer: ‘DNA taxonomy’ — cataloguing barcodes and assigning each to an unnamed species.

Traditional taxonomy cannot keep up with this ‘diversity’. How long will it be until even the specimen is no longer necessary to ‘understand’ the organism?

DNA barcoding generates information, not knowledge. The vast number of barcodes will tell us what we know: life is complex. Museums that encourage DNA barcoding to compete for funding with taxonomy are misguided, as the practice is counterproductive to furthering our understanding of life.

Malte C. Ebach*, Craig Holdrege†

**Buffalo Museum of Science, 1020 Humboldt Parkway, Buffalo, New York 14211, USA*

†The Nature Institute, 20 May Hill Road, Ghent, New York 12075, USA

Don’t mix radiocarbon and calendar years

Sir—Your News Feature “Skeleton keys” (*Nature* **433**, 454–456; 2005) contains two statements that, together, are misleading. First, it is stated that “Clovis people had made it to the southwestern United States by 11,500 years ago”, and second, that Kennewick man is “a 9,000-year-old human skeleton”.

So Kennewick man lived 2,500 years after Clovis? No, in fact, he lived some 4,000 years after (see M. A. Jobling, M. E. Hurler and C. Tyler-Smith *Human Evolutionary Genetics*, Garland Science, New York, 2004).

The confusion arises because the Clovis date is in radiocarbon years, whereas the Kennewick date appears to be somewhere between radiocarbon and regular calendar years: 8,400 radiocarbon or 9,300–9,500 calendar (see www.cr.nps.gov/aad/kennewick/c14memo.htm).

Radiocarbon years differ from calendar years because the amount of ^{14}C in the atmosphere has varied considerably in the past. So the quoted Clovis date of 11,500 years corresponds to approximately 13,500 calendar years.

Please be consistent, and stick to calendar dates so that non-specialists can understand.

Chris Tyler-Smith*, Matthew E. Hurler*, Mark A. Jobling†

**The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK*

†Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK

A global call for new polio vaccines

The end is near but eradication will not be as simple as once thought.

**David L. Heymann,
Roland W. Sutter and
R. Bruce Aylward**

During the next 12 months the world could see the end of the crippling disease polio. That is, if sufficiently high levels of vaccination are achieved in the last few countries with endemic poliovirus, and if surveillance for the paralysis caused by polio manages to find all new cases¹. This will of course require strong political commitment from all concerned. But today there are further concerns that were not anticipated when the World Health Organization (WHO) designed its global eradication strategy in 1988.

As the world nears eradication, the need for new polio vaccines is greater, paradoxically, than at any time in the past 17 years. These vaccines are needed to get to the finish line, to establish a stockpile against future outbreaks, and to maintain the capacity and expertise required to restart oral polio vaccine (OPV) production should the need arise. As with the approach of smallpox eradication in the 1950s, when vaccine research came to a standstill, the research and development of new polio vaccines has slowed. This slowdown must be reversed, and new vaccines once again developed and made available.

Once eradication is achieved, there will be further challenges to face. For example, effective biocontainment of all existing polioviruses is needed to minimize the risk of accidental infections. And OPV use worldwide must be discontinued to minimize the likelihood that mutated strains of the vaccine polioviruses lead to new outbreaks. The WHO must establish an international stockpile of monovalent OPVs or mOPVs (generated from a single strain of virus) and a response mechanism for their use should poliovirus re-emerge after the current trivalent OPV stops being used. Specifically, the world now needs mOPV for type 1, 2 and 3 poliovirus, and a new type of inactivated polio vaccine (IPV) that is manufactured from stocks of weakened live poliovirus strains rather than from wild poliovirus, as is current practice. Finally, it is essential that the expertise required to scale up the production of trivalent OPV (which confers immunity against all three types of



Winning the fight: ensuring that no more children are paralysed by polio will require ongoing vigilance and investment in a vaccine stockpile.

poliovirus) is maintained should the vaccine needs of any future outbreaks exceed the capacity of the stockpile.

Dropping the vaccine

Last year, China had an outbreak of polio, nearly five years after the Western Pacific Region was certified as being clear of polio. This outbreak was not caused by an imported wild poliovirus, but by a circulating vaccine-derived poliovirus (cVDPV) that had mutated from the OPV used to prevent polio. The China outbreak is the fourth cVDPV outbreak to be identified since 2000 (see map, overleaf). The first of these was an outbreak of cVDPV in the Caribbean island of Hispaniola, just six years after the certification of the Americas as polio-free.

In China and in Hispaniola, cVDPVs were detected by national surveillance systems for flaccid paralysis among children younger than 15. Polioviruses are hard to detect because only one in 200 children who are infected develop paralysis. Once a paralysed child is identified, an investigation is conducted and two faecal samples

are collected — at least 24 hours apart, and within 14 days of the onset of paralysis. Faecal samples are then sent to one of nearly 150 WHO-accredited international virology laboratories. Samples are analysed in cell culture for the presence of poliovirus, and subsequent sequencing of the VP1 section of the virus genome is carried out to confirm the detection of cVDPV.

The other two cVDPV outbreaks identified by the WHO laboratories were in the Philippines (2001) and Madagascar (2002). A further outbreak, thought to have begun in Egypt in the 1980s, was identified retrospectively by the testing of stored poliovirus samples. In addition to driving a massive immunization response using OPV, national polio programmes further improved their surveillance systems by collecting faecal specimens from at least 80% of suspected cases, and processing all these specimens in WHO-accredited laboratories.

The current understanding of cVDPVs is that predictable mutation events cause the re-acquisition of virulence and transmission by OPV strains. A total of 31 children have been paralysed in the four recent cVDPV outbreaks, and the outbreaks have been rapidly controlled with the same strategies as those used to eradicate polio worldwide — mass vaccination with OPV and heightened surveillance for paralytic polio. In September 2003, a WHO consultation group on vaccine-derived polioviruses reviewed the epidemiological and genetic sequence data from the known cVDPV outbreaks and concluded that, to prevent future outbreaks after eradication, the use of trivalent OPV for routine immunization must be stopped². An OPV-cessation strategy is therefore being developed for presentation to health ministers at the World Health Assembly in Geneva this May.

Building stockpiles

After the elimination of wild poliovirus transmission worldwide, the risk profile of OPV will change. At that point the well-understood individual risk of paralysis from OPV (roughly one in every 750,000 vaccinations) and the increasingly understood population risk of outbreaks from vaccine-

WHO

derived polioviruses will be greater than the known risk of paralysis caused by wild poliovirus. After eradication, it is thought that many resource-poor countries will stop all polio immunization and focus on other diseases. More industrialized countries are likely to continue vaccinating against polio using IPV.

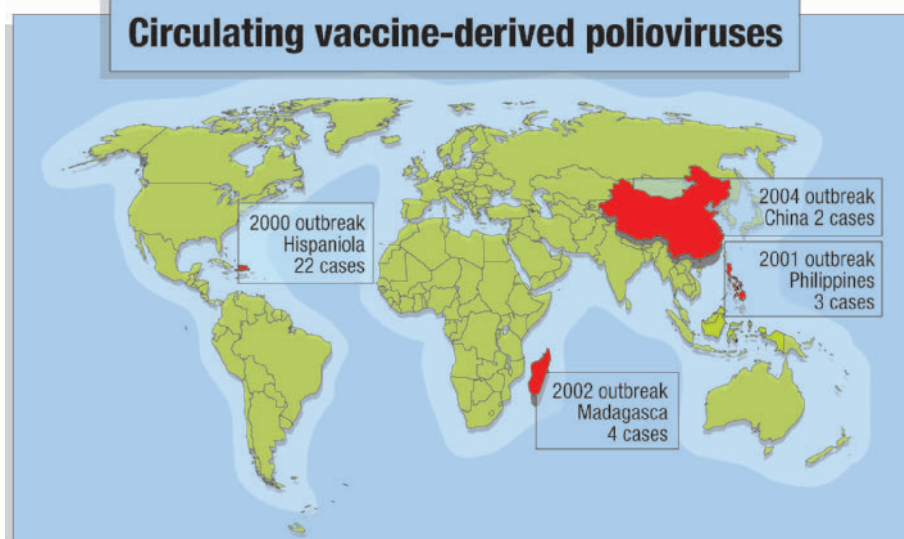
But countries that stop all polio immunization activities can only do so safely if an international vaccine stockpile holds sufficient vaccine to contain future outbreaks³. An international stockpile, with a supply of type 1, 2 and 3 mOPVs, would allow a tailored response to any future outbreaks. Type 1, 2 and 3 mOPVs were licensed and used in the early days of polio control. But as trivalent OPV became the accepted formulation, licences for these vaccines were not renewed. At the request of the WHO, at least two OPV manufacturers have recently begun to again develop and license type 1 mOPV, and licences for this vaccine have been granted in France and India. The development and licensing process for type 2 and 3 monovalents is planned, and must proceed as soon as possible. Once completed, these two new polio vaccines will be available for stockpiling. Type 1 mOPV may also play an important role in breaking the final chains of polio transmission in populous urban areas where only type 1 virus remains, and where a monovalent type 1 vaccine would therefore be more effective.

A combination of modelling exercises and experience with other stockpiles has resulted in a range of suggested sizes for the polio stockpile. Although the final decision has not been reached, the current estimated need is 750 million doses of each type of monovalent vaccine. But the stockpile is planned as a one-time investment. Following the creation of the stockpile, and OPV cessation, the capacity to produce OPV will probably be lost as market incentives to continue production disappear. A potential supply problem could result, should a need for OPV occur that surpasses the original size of the stockpile.

Curbing complacency

Following the interruption of human-to-human transmission of wild poliovirus, the risk of reintroducing virulent poliovirus from a laboratory or a manufacturing facility for IPV will grow as populations lose their immunity. In the early 1990s, a virulent strain used for vaccine manufacturing was isolated, during a routine diagnostic investigation, from a fully vaccinated 19-month-old boy, whose father worked in an IPV manufacturing facility in the Netherlands⁴.

Such laboratory accidents are a real threat to disease eradication. Notably, the most recent known human infections of both smallpox and severe acute respiratory



syndrome (SARS) were due to lapses in laboratory biocontainment procedures⁵. Extra effort is being made to ensure that polio does not follow this trend. As laboratories around the world continue to work with live poliovirus strains, and industrial quantities of virulent poliovirus are produced during vaccine manufacture, the WHO is recommending that all known virulent strains of poliovirus and potentially infectious material (including stool samples) are placed under appropriate levels of biocontainment⁶. By the end of 2004, 152 countries had initiated a survey for wild poliovirus and potentially infectious materials⁷. More than 200,000 facilities have now been searched, and more than 850 facilities have been identified as housing relevant materials. These facilities will either be destroyed or placed under biosafety-level-3/polio-containment conditions.

Production facilities for IPV will pose a particular concern following OPV cessation because they would be the only high-volume and, potentially, high-pressure environments for wild polioviruses. A further measure to decrease the risk of reintroducing wild poliovirus from IPV manufacturing would be the development and licensing of IPV that is formulated from a non-virulent live poliovirus. One such candidate IPV is being developed from Sabin virus strains (Sabin-IPV), but the efficacy and feasibility of large-scale production have yet to be evaluated. Continued research and the development of Sabin-IPV is urgent, not least to help maintain OPV-production expertise.

In summary, the interruption of human-to-human transmission of wild poliovirus worldwide is the first step towards polio eradication. OPV cessation will then be required to guarantee that polio outbreaks do not reoccur. A future polio vaccine stockpile is also essential to ensure the availability of OPV if poliovirus is accidentally reintro-

duced. This will require the licensing of monovalent type 1, 2 and 3 OPVs and their one-time procurement.

Once the stockpile has been established and OPV use universally discontinued, the demand for OPV will decrease, as will the financial incentives to maintain OPV-production capacity and expertise. The licensing and marketing of an effective Sabin-IPV produced from Sabin viruses, rather than from wild poliovirus strains, could help maintain the expertise needed to produce OPV, should future demand be greater than can be met by the stockpile. At the same time, Sabin-IPV could reduce the harmful consequences that would result from an IPV manufacturing accident involving a virulent wild poliovirus.

If the global community wishes to fully achieve full polio eradication, new vaccines are needed — not just to build an OPV stockpile, or to prevent the proliferation of existing poliovirus, but also to help maintain OPV-production expertise, and to decrease the risk of future accidental outbreaks. This is not the polio end-game that was imagined when polio eradication began. But without investment in new vaccines the end may remain beyond reach. ■

David L. Heymann, Roland W. Sutter and R. Bruce Aylward are at the Global Polio Eradication Initiative, World Health Organization, 20 Avenue Appia, CH-1211 Geneva, Switzerland.

1. Heymann, D. L. & Aylward, R. B. *N. Engl. J. Med.* **351**, 1275–1277 (2004).
2. World Health Organization Report of the WHO Consultation on Identification and Management of Vaccine-derived Polioviruses (in the press).
3. Fine, P. E. M., Sutter, R. W. & Orenstein, W. A. in *Progress in Polio Eradication: Vaccine Strategies for the End Game* (ed. Brown, F.) *Developments in Biologicals* Vol. 105 129–147 (Karger, Basel, 2001).
4. Mulders, M. N. et al. *J. Infect. Dis.* **176**, 617–624 (1997).
5. Heymann, D. L., Aylward, R. B., Wolff, C. *Lancet* **363**, 1566–1568 (2004).
6. WHO Global Action Plan for the Laboratory Containment of Wild Polioviruses 2nd edn (World Health Organization, Geneva, 2002).
7. Savolainen, C., Hovi T. *Lancet* **361**, 1187–1188 (2003).

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Emerging physics

A fresh approach to viewing the complexity of the Universe.

A Different Universe: Reinventing Physics from the Bottom Down

by Robert Laughlin

Basic Books: 2005. 304 pp. \$26, £19.99

Philip Anderson

I should make my interests clear right at the start. For many years I have thought that a book such as this should be written, and have been urged to write it myself. I didn't do so, and couldn't possibly have written one as suited as this is for its target audience. *A Different Universe* is a book about what physics really is; it is not only unique, it is an almost indispensable counterbalance to the recent proliferation of books by Brian Greene, Stephen Hawking and their fellows, who promulgate the idea that physics is a science predominantly of deep, quasi-theological speculations about the ultimate nature of things. The enterprise of writing this book has my strong endorsement, then, and any disagreements or criticisms should be read in that light.

The central theme of the book is the triumph of emergence over reductionism: that large objects such as ourselves are the product of principles of organization and of collective behaviour that cannot in any meaningful sense be reduced to the behaviour of our elementary constituents. Large objects are often more constrained by those principles than by what the principles act upon. The underlying laws of physics have no sense of time, give us no clue either to measuring or locating ourselves in space, and provide no clue to identity — we are all made up of nothing but waves in a nonexistent medium

(an analogy that Robert Laughlin draws from Christina Rossetti's poem *Who Has Seen the Wind?*). Our identity and perceptions are all the collective behaviour of 'ghosts', who borrow their reality from each other.

Laughlin gives the reader a quick tour through much of physics (without a single equation). There is a slight emphasis on the quantum theory of condensed matter, in so far as it explains such things as computers (with a sceptical side glance at quantum computation), the properties of ordinary metals, and the like. There is an enlightening discussion of the special quantum phenomena, the Hall and Josephson effects, which through the 'protection' of collective behaviour allow the measurement of Planck's constant and electric charge with enormous accuracy. The term 'protection' was coined by Laughlin and David Pines to express the fact that precise behaviours of large objects can result from, and even benefit from, disorderly behaviour at the atomic level. Some of this will be hard for the layman to follow, but at almost no point is it out of his reach. The pedagogy is leavened by anecdotes, occasional eloquence, and characteristically pungent diction (Laughlin once interrupted a scientific talk with "Liar, liar, pants on fire!").

There are idiosyncratic views of a wide variety of scientific topics. Laughlin reveals his view of nanotechnology with a chapter entitled "Carnival of the baubles" (a view with which one can concur). He gives us some inside information on the Star Wars defence project and its notoriously fraudulent X-ray laser weapon. He continues, comparing unfavourably the 'nanobaubles' of technology

with those of life, the biomolecules, for which he has considerable admiration. Then we hear his own ideas on biology, which will not be to everyone's taste but are certainly thought-provoking. Finally, his view of complexity science surprised and pleased me with its relative benevolence.

Despite the above fulsome praise, this is not by any means a perfect book, even for its purpose. Laughlin is not reliably careful with facts, whether scientific or historical. For example, it has rhetorical value to give his great hero (and winner of two Nobel prizes) John Bardeen mythic status, and to demonize the "engineer" William Shockley, but this is incorrect. Shockley was verifiably contrary and sometimes mean, but he was also a great physicist. It was Shockley, rather than Bardeen, who was responsible for creating the great research centre at Bell Labs with James Fisk and Mervin Kelly. He also hired Bardeen, as well as many of the other stars who graced the place, such as Charles H. Townes, Conyers Herring and Bernd Matthias. Bardeen was human, and was wrong as often as he was right. It would have been instructive to point out that he published two mistaken theories of superconductivity 15 and 7 years before he got the right one. Laughlin's history and emphases are too much those of his generation.

Laughlin makes too much of the role of the renormalization group — a way of averaging out and getting 'universal' properties at the macroscopic level from the messy microscopic details — and other protection principles, as opposed to mechanism, in determining the properties of things. Was it

Pierre Weiss, with his mysterious molecular field and Weiss magnetons, who explained ferromagnetism? Or was it, as I believe, Werner Heisenberg, with quantum theory? Laughlin misleadingly accuses two unnamed physicists of predicting that superconductors be limited to below 30 K (the actual figure was 40 K), when what they said applied specifically to a particular mechanism for which it is true.

In my experience, which incidentally is greater than Laughlin's, underlying causes often enlighten our conceptual thinking as much as precise numbers do, something that Laughlin seems to deny. After condemning astroparticle types for overemphasizing deep thoughts and broad vistas, he seems to reveal a certain measure of 'particle envy' and distaste for the messy, quarrelsome but absorbing ways of doing the sciences in the real world. What made Bardeen great, as indeed he was, was his stubbornness and experimental taste, and Laughlin dismisses these values.

Those who devour the work of Greene, or decorate their coffee table with Hawking, will find this book a useful antidote. It should spike the interest of those who read the physics popularizers, although in its personalized coverage and opinionated style it is *sui generis*. My message is this: buy the book. ■

Philip Anderson is in the Condensed Matter Physics Group, Department of Physics, Princeton University, Princeton, New Jersey 08544, USA.

Don't talk to the animals

Doctor Dolittle's Delusion: Animals and the Uniqueness of Human Language

by Stephen R. Anderson
Yale University Press: 2004. 368 pp.
\$35, £22.50

Neil Smith

Doctor Dolittle was the hero of a series of children's books written by Hugh Lofting (1886–1947). The doctor's ability to talk to every animal in its own language had a seductive appeal that finds current expression in the widespread belief that the communication systems of animals, from bees to bonobos, are essentially similar to human language. So we should be able to learn their languages or, just as good, they could learn ours. Stephen Anderson pours cold water on this belief, arguing convincingly that it is a delusion.

He provides a masterly overview of what is currently known about the communicative abilities of a wide range of creatures: the dance of honeybees, the communicative croaking of frogs, the warning cries of monkeys, and the remarkable cognitive abilities of bonobos and parrots. Much of this is

superficially familiar from other popular accounts, but Anderson's synthesis provides illuminating comparisons with the infinitely more sophisticated resources of human language, whether spoken or signed. There are undeniable parallels between humans and other animals, but the differences are equally striking and confirm the view that our language is qualitatively different from theirs.

Bees famously indicate the direction, distance and quality of sources of pollen to their fellow workers by means of a 'waggle dance', which is often taken to show that they have a 'cognitive map' of the local terrain. In a meticulous dissection of the properties of this dance, Anderson undermines this claim, showing that the bees' perception of distance is largely a function of differences in their visual experience. When bees are made to fly through tunnels with visual patterns on the walls, the distance they indicate corresponds to the complexity of the pattern to which they have been exposed. The dance, then, reflects the bees' subjective experience, rather than a map of the external world.

We are evolutionarily rather remote from bees; closer parallels to human language and how we learn it can be found in birds. Current views of first-language acquisition in children treat it as a process of selection, rather than instruction. Humans are born with a set of principles known as 'universal grammar' that define the notion of possible

Exhibition

Hirst's hobbit

The maxim that the best science possesses an artistic grace has probably never been applied to the down-and-dirty world of palaeontology. First there's all the digging, and then there's the arguing over how the fragmentary findings should be interpreted and slotted into the big picture. But a new painting by Damien Hirst, *A New and Diminutive Species of Human Being Has Been Discovered*, brings the face of *Homo floresiensis* (the 'hobbit') — one of palaeontology's most iconic recent images — to the gallery.

Photorealism — the painstakingly faithful reproduction of photographic images on canvas — is a surprising new direction for Hirst, arguably the doyen of the laddish Britart scene of the 1990s. Better known for pickling animals inside glass boxes, he has often engaged with science before. One of his trademark 'dot paintings' travelled to Mars aboard the ill-fated Beagle 2 lander and, given a smoother landing, would have been used to calibrate the craft's onboard cameras.

Hirst's other photorealist creations include images of the Iraq conflict, vivisection and a haunting reproduction of a British police anti-drugs campaign poster featuring the gaunt features of a now-dead crack addict. But what inspired him to take on *H. floresiensis*,



unveiled in *Nature* last October? "It's just an excuse to paint skulls," he says. Perhaps, but the hobbit's discoverers may nonetheless be amused to see the fruit of their labours raised to an art form.

A New and Diminutive Species of Human Being Has Been Discovered is part of the exhibition 'The Elusive Truth', which can be seen at the Gagosian Gallery in New York until 23 April.

Michael Hopkin

human language and a set of parameters that characterize possible variation among languages. The acquisition of language consists largely of fixing the values of these parameters on the basis of clues in the input. Importantly, this theory predicts that there are logically possible, but linguistically impossible, mistakes that children cannot make.

The parallel with birdsong is striking. Many birds have songs that develop appropriately only after interaction with conspecifics: the song is partly innate and partly dependent on experience. Nightingales, like children, only make mistakes that correspond to patterns that could occur as possible song elements for their species, just as children only make mistakes that are licensed by universal grammar.

We are evolutionarily closer to birds than to bees, but we are closer still to other primates. Their systems of communication, however, are less similar to human language than is popularly supposed. Vervet monkeys have distinct alarm calls for leopards, eagles and snakes. These calls can be extended to new types of threat — humans, for instance — and they are under some degree of voluntary control, yet they do not seem to ‘refer’ to the respective animals in the way we refer with our language. The calls can affect behaviour but not knowledge. Similarly, attempts to teach American Sign Language to chimpanzees have made it clear that, although human infants read intentions into the actions of others, chimps never do.



The grizzly truth? Only humans use syntax, so sharing a language with animals is left to films such as *Doctor Dolittle*.

To complement his critique of ‘animal language’, Anderson also outlines what is special about human language: in a word, syntax. Our vocabularies are dramatically larger than those of other animals, and our sound systems are more complex, but the essential design property of human language is syntax — the way we use combinations of words to convey meaning. This concept is alien to the communication systems of other species.

What animals learn is impressive and their cognitive abilities may be remarkable. But they never master anything like a human language and seem incapable of doing so: the complexity of their grammar is not remotely comparable to ours. This complexity is exemplified at length in the book, but two examples should suffice. First, the essence of syntax is recursion: the possibility of including one sentence inside another ad infinitum. For example: [Anderson discusses

the claim that [many people think that [animals can talk]]]. Second, we all have subtle and consistent intuitions not only of what is possible in our language, but also of what is impossible. In [John expects to visit him], “John” and “him” must refer to different people, but consider this: [I wonder who [John expects to visit him]]. Here, “John” and “him” can, but need not, refer to the same individual. Judgements such as these have no parallel in the communication systems of other animals.

Anderson’s elegant book contains a host of other insights and observations. He concludes that, just as the dance of bees, the song of birds and the calls of monkeys are unique to their respective species, so human language is unique to us. ■

Neil Smith is in the Department of Phonetics and Linguistics, University College London, Gower Street, London WC1E 6BT, UK.

Aventis Science Book Prize shortlist announced

The Royal Society has announced the shortlist for this year’s Aventis Prizes for Science Books’ General Prize, which celebrates the very best in popular science writing for adults.

Critical Mass by Philip Ball (William Heineman) takes a look at the application of physics to the collective behaviour of society. “This book is impressively clear and breathtaking in scope... For anyone who would like to learn about the intellectual ferment at the surprising junction of physics and social science, *Critical Mass* is the place to start.” Steven Strogatz (*Nature* **428**, 367–368; 2004).

Robert Winston’s **The Human Mind** (Bantam Press/Transworld) is an examination of the workings of our brains for an adult audience, inspired by his recent television series. His book *What Makes Me, Me?* (Dorling Kindersley),

which takes a wider view of how the human body functions for a younger audience, has been shortlisted for the Junior prize.

In **The Ancestor’s Tale** (Weidenfeld & Nicolson), Richard Dawkins “views species as pilgrims marching into the past, joining each other genetically on a 3-billion-year journey to evolution’s Canterbury: the first ‘replicator’”. Jerry Coyne (*Nature* **431**, 903–904; 2004).

In **Matters of Substance** (Penguin, Allen Lane), Griffith Edwards presents a lucid account of drug use and control, taking the radical view that the effect of any drug is just as dependent on the social, historical and psychological context as on its chemical structure.

In **The Earth** (HarperCollins), Richard Fortey “offers a clear, graphic and entertaining

exposition of the manner in which, over an eon, the observed geological phenomena have achieved their present state. And he forcefully reminds us that events remotely embedded in deep time may yet be highly relevant as determinants for the lifestyles of modern human communities.” Gordon L. Herries Davies (*Nature* **428**, 697–698; 2004).

A review of **Why Life Speeds Up As You Get Older** by Douwe Draaisma (Cambridge University Press), an examination of the nature of memory, will appear in next week’s issue of *Nature*.

The General Prize judging panel consists of author Bill Bryson, who won in 2004, weather forecaster Lisa Burke, Sian Ede, who is a renowned authority on art and science interactions, neurophysiologist Mark Lythgoe, and poet Ruth Padel. The winners will be announced on 12 May 2005.

Stirring the primordial soup

RNA world: does changing the direction of replication make RNA life viable?

William R. Taylor

It is now widely believed that almost four billion years ago, before the first living cells, life consisted of assemblies of self-reproducing macromolecules. The molecular candidate thought to mediate this activity was RNA, which can combine the necessary properties of encoding information and catalysing chemical reactions — functions that are now fulfilled largely by DNA and proteins, respectively. From theoretical arguments, it can be expected that a system of interacting molecules will give rise to complex, and even life-like, behaviour, but there is still debate about whether RNA was the first or the only macromolecule to participate in such activity, with both protein and DNA (or any combination with or without RNA) representing alternatives.

Circumstantial evidence for the central position of RNA in the origin of life can be found in 'relic' pieces of RNA that hold a few of the most important functions in the cell. Perhaps the most convincing observation is that, in the synthesis of proteins on the ribosome, the key chemical event — peptide-bond formation — is catalysed solely by RNA, suggesting that primacy lies with RNA rather than protein. A major impediment to full acceptance of an 'RNA world' is that, although it can easily be imagined that a pure RNA machine (a proto-ribosome) can make proteins, there is no equivalent RNA machine to make RNA (a ribopolymerase). All the RNA we know is made by protein, leading to perhaps the original 'chicken-and-egg' problem of which came first.

Some mechanisms for replication in the RNA world have been put forward, and following the current systems of protein polynucleotide synthesis, all involve the creation of a complementary daughter strand using Watson–Crick base-pairing. But from a mechanistic viewpoint, such a model contains a fundamental problem: if a ribopolymerase were to make a complementary copy of itself, it would need to recopy this to obtain a new functional ribopolymerase. This implies that both the ribopolymerase sequence and its complement would have to coexist. But if these two copies came together, the result would be a double stranded Watson–Crick helix (as found in some RNA viruses) — not a new ribopolymerase. Even if both sequences had well determined secondary structures, the perfect complementarity of the Watson–Crick pairing would act as a sink, leading to a sterile population of double-stranded molecules.

In a world without any other type of



molecule (such as protein) to prevent these unwanted interactions, it might be concluded that a pure RNA world could not have been viable. But what if the ribopolymerase did not synthesize a complementary strand? From a chemical viewpoint, there is no reason why a polymerase must make a complementary strand that runs in the reverse direction to the template strand. In modern protein polymerases, nucleotide triphosphates are added to the 3' end of the transcript without the direct participation of the template strand. If the template strand was flipped (making a parallel complement), then all that would be lost is some capacity for the template and transcript to remain base-paired, as parallel nucleic acid strands cannot form a duplex with Watson–Crick base-pairing. In an RNA world, the loss of this interaction would be an advantage — preventing the formation of a dead-end double helix.

Starting replication at the 3' end of the template strand, a transcript cannot be recopied until it is completed. In an RNA world, this strategy would leave a full-length transcript exposed to a hostile environment in which it might make many spurious interactions with other RNA molecules (especially its own complement) and, unless it quickly adopted a compact conformation, it would

be susceptible to hydrolysis. By contrast, a polymerase that begins at the 5' end produces a transcript on which retranscription can start almost immediately, minimizing the exposure of single-stranded RNA and avoiding hybridization. This strategy is similar to the immediate translation of messenger RNA in bacteria, which allows thermophilic species to minimize the exposure time of the single-stranded message at high temperatures. Immediate retranscription would also confer a similar protection in a hotter primeval world. Two ribopolymerases operating in tandem as a dimer could take a ribopolymerase sequence as a template and produce a new ribopolymerase. However, the intermediate parallel-complementary strand need not be discarded, and could be picked up by another ribopolymerase, potentially leading to a vast network of linked replication.

A difference in transcription direction may also explain why the RNA world eventually had to become extinct. When life progressed to a compartmentalized, genome-based system, it would be necessary to have a break in transcription to allow the physical separation of the messages. If each new copy were to be immediately retranscribed, a continual cascade of transcription would ensue. Although this may be efficient in a 'soup', it gives no defined point for the separation of individual genomes. If proteins emerged alongside such a system adopting the 'parasitic' role suggested by Freeman Dyson, those that gained some capacity to synthesize a complementary strand would not only create genomes that could be compartmentalized into cells, but the complementary strands would hybridize with the ribopolymerases, hastening their decline. The origin of protein polymerases may therefore have driven the ascent of cellular genomes and the eventual end of the RNA world. ■

William R. Taylor is in the Division of Mathematical Biology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK.

FURTHER READING

- Joyce, G. F. *Nature* **338**, 217–224 (1989).
 Gesteland, R. F., Cech, T. R. & Atkins, J. F. (eds) *The RNA World* (Cold Spring Harbor Lab. Press, 1999).
 Maynard Smith, J. & Szathmáry, E. *The Major Transitions in Evolution* (Oxford Univ. Press, 1995).
 Taylor, W. R. *Comp. Biol. Chem.* **28**, 313–319 (2004).
 Dyson, F. *Origins of Life* (Cambridge Univ. Press, 1985).

Acknowledgements

Bob Cox and Steve Smerdon are thanked for valuable discussion.

A planet that blinks

Karl Stapelfeldt

Infrared radiation from two extrasolar planets has been measured from the dip in total light as the planets pass behind their parent stars — a milestone on the road to the direct imaging of such planets.

Detecting light from planets orbiting stars other than the Sun is one of the most challenging tasks in observational astronomy: a trickle of planetary photons must somehow be separated from the flood of light emitted by the planet's parent star. The classic approach is to employ one or more large telescopes, suppress the stellar glare as much as possible by means of coronagraphic masks or destructive interference, and extract the planetary signal from the residual starlight.

In a few cases at least, another approach is possible. Two groups, Deming *et al.*¹ (page 740 of this issue) and Charbonneau *et al.*², now report the first measurements of infrared light from two different planets circling Sun-like stars. The measurements were made using a relatively small telescope (NASA's 85-cm-aperture Spitzer Space Telescope; Fig. 1), and without any specialized starlight-suppressing instrumentation. These somewhat surprising results were only possible because each body is a 'transiting planet' — one whose orbital motion causes it alternately to eclipse and be eclipsed by its parent star. During the latter case (a 'secondary eclipse'), the planet's light blinks out and the drop in the combined light of the system is equal to the light emitted by the planet. In a clever reversal, it is therefore the suppression of the planet's light, not the star's, that allows the planetary and stellar emissions to be distinguished.

Most of the 152 known extrasolar planets³ were discovered through precision measurements of small stellar accelerations induced by the gravity of the orbiting planet. The exceptions are planets discovered through transit surveys, where a recurring decrease of around 1% in stellar brightness indicates the repeated passage of a Jupiter-sized planet in front of the star. Systems with such transits ('primary eclipses') are particularly valuable because they are the only cases where a planet's mass and radius can be determined accurately. That these systems also offer the potential to measure planetary emissions through their secondary eclipses was noted when the first transiting planet was discovered five years ago⁴.

The known transiting planets all orbit close to their parent stars at radii of about 0.04 astronomical units (AU; for comparison, Mercury is about 0.4 AU away from our own Sun). Their expected equilibrium tem-

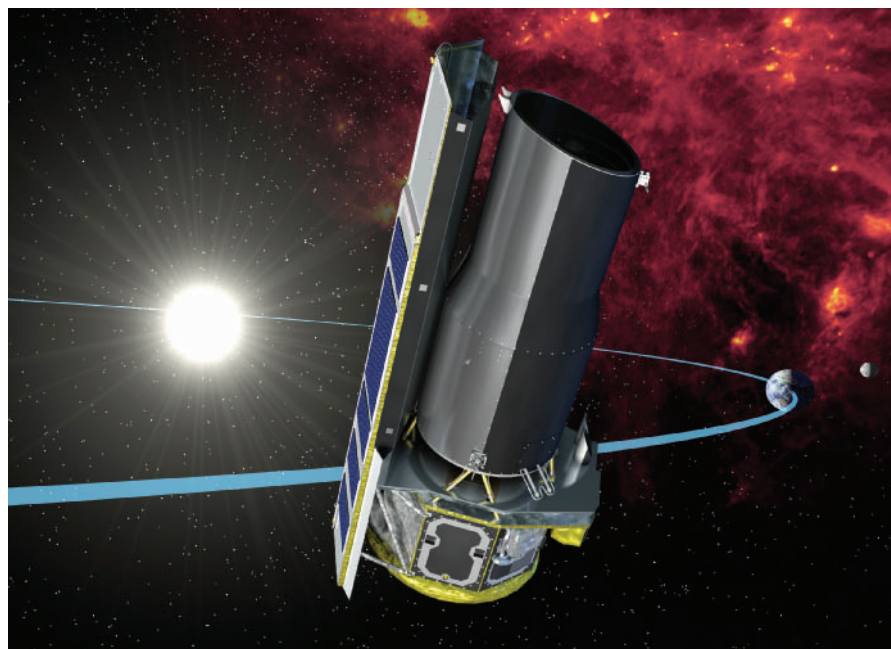


Figure 1 Planetary spy. An artist's impression of the Spitzer Space Telescope. The largest space-based infrared telescope, with an aperture of 85 cm, Spitzer trails the Earth in a heliocentric orbit.

peratures are near 1,000 K and so their emissions should peak at infrared wavelengths. Attempts to detect the infrared signature of the secondary eclipse for the first transiting exoplanet to be discovered, HD 209458b, have been unsuccessful using large, ground-based telescopes⁵. The high sensitivity and stability of the Spitzer telescope enabled the direct detection of infrared emissions from both HD 209458b (ref. 1) and another transiting planet dubbed TrES-1 (ref. 2). Each system has a previously discovered giant planet with an orbital period of about three days. Neither star is particularly close to our system, lying at distances of 150 and 500 light years respectively.

Dips in each star system's brightness during the secondary eclipses were measured by the Spitzer telescope in one or two broad spectral bands. Both planets appear to be about 1/400 as bright as their parent star at mid-infrared wavelengths. The corresponding temperature found for these planets' star-facing hemispheres is about 1,100 K — in the range of values expected for planets in thermal equilibrium that are orbiting close to their central stars. Furthermore, the timing of the secondary eclipses constrains each planet's 'orbital eccentricity' (a number that expresses the deviation of an orbit from the

circular, with a value of zero indicating a perfect circle). If the orbit is circular, primary and secondary eclipses should alternately recur at intervals of half an orbital period, and this is what was observed. It is therefore unlikely that the orbit of either planet is particularly eccentric.

This result seems to eliminate one possible explanation for the large radius of HD 209458b, the planet investigated by Deming *et al.*¹, which has been shown by primary eclipse measurements to be 35% larger than that of Jupiter. Flexing of the planet's interior by tidal forces, and the concomitant dissipation of heat, could maintain the planet's large radius against gravitational compression⁶. However, this hypothesis requires a small but significant eccentricity in the planet's orbit, and the new Spitzer results render this unlikely. An alternative explanation — that the planet's atmosphere is highly inflated by the scorching heat so close to its parent star — would account for the large radius of HD 209458b (ref. 7) but not the smaller, Jupiter-like radii of TrES-1 and other transiting planets. New theories may have to be sought.

These results are noteworthy as the first detections of light emitted by extrasolar planets of Sun-like stars, and will surely be expanded on in future observations with the

NASA/JPL

Spitzer telescope. Such observations might include measurements in several spectral bands to constrain better each planet's temperature; observations near primary eclipse to measure the temperature of the planet's night side and derive the day–night temperature contrast as an indicator of the efficiency of any atmospheric circulation; and spectroscopy during secondary eclipse in the hope of revealing the atmospheric composition. The diversity of planetary properties might be explored by targeting any bright, newly discovered transiting planets with this full set of investigations. Much progress can be made during the next three-and-a-half years, before Spitzer's science mission is curtailed by the exhaustion of its liquid helium coolant.

As interesting as these results^{1,2} are, they can only be realized for a small set of hot, Jupiter-like planets that transit their parent stars. To directly detect and characterize planets outside our Solar System over their full range of sizes, temperatures, orbital radii, orbital inclinations and atmospheric compositions⁸ will require the development of ultra-high-contrast imaging systems on large telescopes. The detection of a probable planetary companion at a distance of 55 AU from the very young brown dwarf 2MASSWJ1207334-393254 (refs 9, 10), although it is not technically difficult at a contrast of only 100:1, exemplifies the spatially resolved imaging studies that will be needed to study the whole range of planetary systems.

Adaptive optical systems are in development that will reach contrasts greater than 10^6 using today's large, ground-based telescopes. Laboratory tests of precision adaptive optics coupled to a coronagraph in a vacuum have now demonstrated imaging at 10^9 contrasts in close proximity to an artificial stellar source¹¹. These tests point the way towards future high-contrast space telescopes such as Eclipse and the Terrestrial Planet Finder, which are still in the planning stages. The adventure to be found in photons from planets outside the Solar System is just beginning.

Karl Stapelfeldt is in the Astrophysics Element, Earth and Space Sciences Division, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA.
e-mail: krs@exoplanet.jpl.nasa.gov

1. Deming, D., Seager, S., Richardson, L. J. & Harrington, J. *Nature* **434**, 740–743 (2005).
2. Charbonneau, D. *Astrophys. J. Lett.* (submitted); preprint available at <http://arxiv.org/abs/astro-ph/0503457>.
3. Schneider, J. *The Extrasolar Planets Encyclopaedia* www.obspm.fr/encycl/encycl.html (2005).
4. Charbonneau, D. *et al. Astrophys. J. Lett.* **529**, L45–L48 (2000).
5. Richardson, L. *et al. Astrophys. J.* **584**, 1053–1062 (2003).
6. Bodenheimer, P., Lin, D. N. C. & Mardling, R. A. *Astrophys. J.* **548**, 466–472 (2001).
7. Burrows, A., Sudarsky, D. & Hubbard, W. B. *Astrophys. J.* **594**, 545–551 (2003).
8. Burrows, A. *Nature* **433**, 261–268 (2005).
9. Chauvin, G. *et al. Astron. Astrophys.* **425**, L29–L32 (2004).
10. Schneider, G. *et al. Am. Astron. Soc. 205th Meeting*, 11.14 (2004).
11. Trauger, J. *et al. Proc. SPIE* **5487**, 1330–1336 (2004).

Pharmacology

Marijuana and your heart

Michael D. Roth

Marijuana smoke can have harmful effects on the heart. But one of its active components may ease inflammation and slow the progression of coronary artery disease.

The discovery of cell-surface receptors that bind to the major active component of marijuana, Δ^9 -tetrahydrocannabinol (THC), has led to an explosion of research into the biological properties of marijuana and cannabinoids¹. THC binds with equal affinity to two different receptors — CB1 and CB2. CB1 is present at high levels on brain cells and at much lower levels on cells outside the nervous system. By contrast, CB2 receptors occur exclusively on cells outside the nervous system and in particular on cells of the immune system. This pattern of distribution suggests different functions for CB1 and CB2 that might be exploited therapeutically.

On page 782 of this issue, Steffens *et al.*² evaluate whether THC can protect against the development of atherosclerosis, a disease

in which a combination of fatty deposits and inflammation leads to 'plaques' that obstruct coronary arteries, causing angina and heart attacks. The authors suggest that the immunosuppressive properties of THC, and specifically those mediated by CB2 receptors, might be developed to treat heart disease.

The capacity of cannabinoids to regulate immune function is now well established. Exposing immune cells to THC alters their ability to produce certain signalling proteins called cytokines. When THC is administered to animals *in vivo* and to human cells *in vitro*, it suppresses the production of protective cytokines and increases the production of immunosuppressive cytokines. As a result, mice treated with THC fail to develop protective immunity against opportunistic

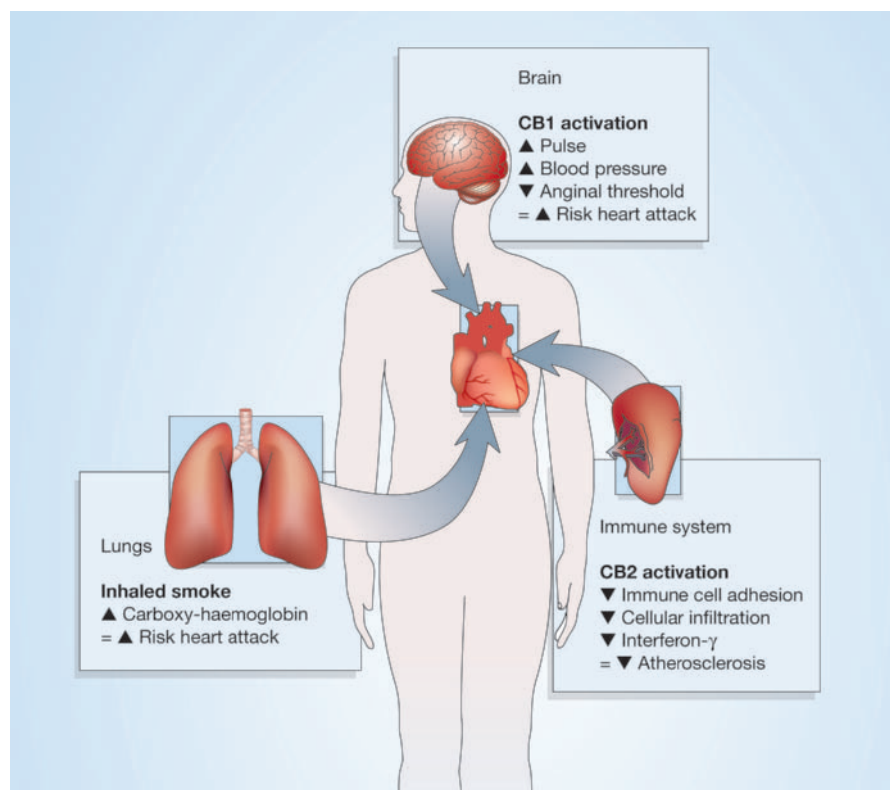


Figure 1 The conflicting effects of marijuana smoke on atherosclerotic heart disease. Simultaneous activation of cannabinoid receptors CB1 and CB2 by the major active component of marijuana, Δ^9 -tetrahydrocannabinol (THC), may counteract the protective effect of THC reported by Steffens *et al.*². Although the isolated effects of THC on CB2 receptors might reduce inflammation and the infiltration of immune cells into atherosclerotic plaques, simultaneous activation of CB1 receptors in the brain induces a cardiovascular stress response that can significantly increase cardiac oxygen consumption and reduce blood flow in coronary arteries. In addition, marijuana smoking increases carboxy-haemoglobin levels, reducing the capacity of the blood to carry oxygen. Smoking marijuana may therefore have a negative overall impact on atherosclerotic heart disease.

infections and cancer³. Similarly, immune cells collected from the lungs of marijuana smokers produce lower than normal amounts of several cytokines and fail to produce nitric oxide (another intermediary in the immune system), severely limiting their ability to kill bacteria⁴.

Steffens and colleagues² set the stage for their work by demonstrating that immune cells expressing CB2 receptors infiltrate atherosclerotic plaques in humans and in a strain of mice that is used to study atherosclerosis (ApoE^{-/-} mice). In this mouse model, the animals develop progressive narrowing of their arteries as lipids and inflammatory cells called macrophages enter the walls of their blood vessels and produce plaques⁵. When low doses of THC (1 mg per kg body weight per day) were added to their diet, the progression of atherosclerosis was markedly slowed. Mice that were fed THC still had elevated levels of serum lipids but had fewer plaque-infiltrating macrophages when compared with controls, suggesting an effect on immune function.

The authors go on to substantiate the immunosuppressive properties of THC on the migration, infiltration and function of immune cells in this model. Spleen cells collected from THC-treated mice responded poorly to stimulation *in vitro*, showing limited proliferation and impaired production of interferon- γ , a cytokine involved in atherosclerosis. Macrophages harvested from THC-treated animals expressed little of the messenger RNA encoding the CCR2 receptor protein, and were poorly responsive to the CCR2 ligand, MCP-1. Both CCR2 and MCP-1 are involved in the migration of macrophages and have roles in atherosclerosis⁶. These effects were blocked when mice were pre-treated with a selective antagonist to the CB2 receptor or when macrophages were collected from mice that lack functional CB2 receptors.

Using live vascular microscopy, the authors observed that adhesion of immune cells to the internal surface of blood vessels was considerably reduced in mice that were fed THC. Again, this effect was blocked by pre-treating the animals with a CB2 antagonist. The protective effects of THC occurred at very low doses, producing blood THC levels well below the range usually associated with activation of CB1 receptors in the brain. The authors conclude that low doses of THC, or perhaps selective CB2 ligands, should be further investigated for their possible use in treating human atherosclerosis.

The findings by Steffens *et al.* are striking, but they should not be taken to mean that smoking marijuana is beneficial for the heart. The dose-response curve to THC in this study was very narrow and U-shaped, with higher and lower concentrations failing to produce protective effects. It would be difficult to achieve such specific concentrations

in the blood by smoking marijuana. Also, no studies have been performed in humans to evaluate the effects of THC on atherosclerosis. As the authors note, the ApoE^{-/-} mice develop extremely high levels of serum lipids, and THC, which is very fat-soluble, is likely to be stored at high local concentrations within atherosclerotic lesions⁷. Whether this local storage occurs, and whether the same effect will occur in human atherosclerotic lesions, remains to be determined.

Finally, THC binds to and activates CB1 and CB2 receptors with similar affinity. Marijuana smoking, acting through its effect on CB1 receptors in the brain, increases the pulse rate, produces an acute rise in blood pressure and then results in sudden falls in blood pressure upon standing or walking (Fig. 1). These effects lower the exercise threshold for angina, and are an independent risk factor for heart attack and stroke^{8,9}. When inhaled, marijuana smoke also increases the concentration of carboxy-

haemoglobin in the blood, impairing oxygen delivery. Ultimately, to take advantage of the positive effects reported by Steffens *et al.* will probably mean developing cannabinoids that target CB2 receptors, rather than using marijuana or oral THC as medicines. ■

Michael D. Roth is in the Division of Pulmonary and Critical Care, Department of Medicine, CHS 37-131, David Geffen School of Medicine, University of California, Los Angeles, California 90095-1690, USA.
e-mail: mroth@mednet.ucla.edu

1. Di Marzo, V., Bifulco, M. & De Petrocellis, L. *Nature Rev. Drug Discov.* **3**, 771–784 (2004).
2. Steffens, S. *et al. Nature* **434**, 782–786 (2005).
3. Klein, T. W. *et al. J. Leukoc. Biol.* **74**, 486–496 (2003).
4. Shay, A. H. *et al. J. Infect. Dis.* **187**, 700–704 (2003).
5. Meir, K. S. & Leitersdorf, E. *Arterioscler. Thromb. Vasc. Biol.* **24**, 1006–1014 (2004).
6. Charo, I. F. & Taubman, M. B. *Circ. Res.* **95**, 858–866 (2004).
7. Nahas, G. G., Frick, H. C., Lattimer, J. K., Latour, C. & Harvey, D. *Hum. Psychopharmacol.* **17**, 103–113 (2002).
8. Mittleman, M. A., Lewis, R. A., Maclure, M., Sherwood, J. B. & Muller, J. E. *Circulation* **103**, 2805–2809 (2001).
9. Geller, T., Loftis, L. & Brink, D. S. *Pediatrics* **113**, e365–e370 (2004).

Tuberculosis

The genetics of vulnerability

Nada Jabado and Philippe Gros

Susceptibility to tuberculosis is known to be under complex genetic control in humans, but what are the genes involved? A mouse strain that is unusually prone to the disease shows the way.

It is estimated that as much as one-third of the world's population is infected with the tubercle bacillus *Mycobacterium tuberculosis*. Yet only one in ten infected people actually develops active tuberculosis (TB), suggesting that innate mechanisms of immune defence (among other factors) can often contain the infection¹. Innate susceptibility to TB has been known for years to be genetically controlled, but this genetic component is very complex and has been difficult to unravel². On page 767 of this issue, however, Pan *et al.*³ report that they have identified a host gene that controls susceptibility to TB in mice. This gene is expressed in macrophages — the host cells in which *M. tuberculosis* replicates — and appears to determine the type of death that macrophages will suffer following infection. These studies not only provide insight into a novel aspect of TB pathogenesis, but, if validated in humans, may reveal new molecular targets for drug treatments.

Tuberculosis is endemic in many parts of the world and continues to be a major global health problem, resulting in 2 million deaths a year. Contributing factors include poor health services, widespread poverty and other socioeconomic problems, the HIV epidemic and the appearance of multidrug-resistant *M. tuberculosis*^{1,4}. The bacterium (Fig. 1, overleaf) infects lung macrophages,

and is concentrated in granulomas — solid clumps of macrophages, lymphocytes (types of immune cell) and epithelial cells. These granulomas appear to contain the infection physically. TB can remain dormant in this form for years. But, in a proportion of individuals, the protective granulomas or other aspects of host immunity break down, leading to active pulmonary TB and the spread of the disease through the production of aerosols by the lungs.

The host factors that underlie protective immunity and that may fail in active TB are poorly understood, but they are at the centre of TB pathogenesis⁵. Studies of twins, and population studies in regions in which TB is endemic (as well as during 'first-contact' epidemics and local outbreaks), indicate that host genetics plays a part. But this component involves numerous genes, with other factors — such as genetic variability and gene-environment interactions — further complicating the identification and analysis of single-gene effects in humans^{2,6}.

In their approach to the problem, Pan *et al.*³ have studied the inbred C3HeB/FeJ mouse strain. In these mice, extreme susceptibility to TB is associated with uncontrolled replication of *M. tuberculosis* in the lungs, with animals rapidly succumbing to infection. This susceptibility is controlled in part by a chromosomal region, or locus, called



Figure 1 Agent of tuberculosis: the tubercle bacillus, *Mycobacterium tuberculosis*.

sst1 (for 'supersusceptibility to tuberculosis 1') that the authors previously mapped⁷ to chromosome 1. Now, Pan *et al.* have generated a mouse strain in which the *sst1* locus from the C3HeB/FeJ mice has been replaced with the same region from C57BL/6J mice, which are resistant to TB. The authors used this 'congenic' strain to study the contribution of *sst1* to the pathogenesis of TB.

Pan *et al.* first show that the acquisition of resistance in the congenic strain is associated with an increased capacity of bone-marrow-derived macrophages to restrict intracellular replication of *M. tuberculosis*. Interestingly, resistance is also linked to the induction of apoptosis (a strictly regulated form of cell suicide) in macrophages following infection. Macrophages in susceptible mice, by contrast, die by necrosis — an uncontrolled process that may be less likely to restrict bacterial spread.

Scrutiny of the *sst1* chromosomal region for candidate genes that are expressed in macrophages then led the authors to a gene that they call *Ipr1* (for 'intracellular pathogen resistance 1'). *Ipr1* is expressed in resistant macrophages and is induced upon *M. tuberculosis* infection, but is absent from susceptible cells. The authors' gene hunt was complicated by the fact that the *sst1* locus includes part of a 'homogeneously staining region' — a large, unstable region of repeated DNA that contains several rearranged copies of *Ipr1*-related sequences. This may explain the susceptibility of C3HeB/FeJ mice: the instability in this region may have decapitated the regulatory sequences of the *Ipr1* gene, preventing it from being expressed.

Pan *et al.* went on to validate their findings by showing that reintroducing full-length *Ipr1* into C3HeB/FeJ mice could partially suppress the replication of *M. tuberculosis* in the lungs *in vivo*, and in macrophages infected *in vitro*. A particularly

exciting finding is that *Ipr1* expression in such genetically altered C3HeB/FeJ macrophages could also restrict the replication of another bacterium, *Listeria monocytogenes* — suggesting a general role for *Ipr1* in innate macrophage defences against intracellular infections.

So what exactly does *Ipr1* do? The gene encodes a protein known as Ifi75, which contains several sequence motifs that indicate that it is localized to the cell nucleus and has a role in regulating gene expression. In support of that idea, Ifi75 is a relative of the human protein SP110 (ref. 8) — a proposed regulator of gene transcription. Like *Ipr1*, SP110 is regulated by interferons, signalling molecules involved in immunity. SP110 also interacts with certain viral proteins, including proteins from hepatitis C virus and Epstein–Barr virus⁹. All of this places *Ipr1* at the interface of host–pathogen interactions, possibly participating in transcriptional activation in macrophages in response to intracellular pathogens. (See also the supplementary information in ref. 3.)

It will be interesting to see how the presence of products from bacteria as different as *Listeria* and *Mycobacterium* can activate *Ipr1*, and what other host proteins may be involved in this signalling. Other questions are: which genes are in turn activated by *Ipr1*, and how do they contribute to the resistance of macrophages to bacterial replication in general, and to the induction of apoptosis in particular? Might this response involve the activation of 'inflammatory' or 'apoptotic' caspase enzymes? And does this pathway run parallel to, or intersect with, other pathogen-sensing pathways, such as those triggered by extracellular pathogens through the so-called Toll-like receptors or other proteins, including those of the NBS-LRR (nucleotide-binding site leucine-rich repeat) or NOD (nucleotide oligomerization domain) families¹⁰?

Suffice it to say that, once again, careful genetic studies in the laboratory mouse have delivered an unexpected gift. More exciting biology in an area of immense interest for global health is sure to follow. ■

Nada Jabado is in the Department of Pediatrics, Montreal Children's Hospital, McGill University Health Center, Montreal, H3Z 2Z3, Canada.

Philippe Gros is in the Department of Biochemistry, McGill University, Montreal, H3G 1Y6, Canada.

e-mail: philippe.gros@staff.mcgill.ca

1. Ravighione, M. C. *Tuberculosis* **83**, 4–14 (2003).
2. Bellamy, R. *Genes Immun.* **4**, 4–11 (2003).
3. Pan, H. *et al.* *Nature* **434**, 767–772 (2005).
4. Bloom, B. R. *N. Engl. J. Med.* **346**, 143–145 (2002).
5. Smith, I. *Clin. Microbiol. Rev.* **16**, 463–496 (2003).
6. Casanova, J. L. & Abel, L. *Annu. Rev. Immunol.* **20**, 581–620 (2002).
7. Kramnik, I. *et al.* *Proc. Natl Acad. Sci. USA* **97**, 8560–8565 (2000).
8. Kaderit, S. *et al.* *J. Biol. Chem.* **268**, 24432–24441 (1993).
9. Nicewonger, J. *et al.* *J. Virol.* **78**, 9412–9422 (2004).
10. Inohara, N. & Nunez, G. *Nature Rev. Immunol.* **3**, 371–382 (2003).



100 YEARS AGO

A Study of Recent Earthquakes. A subject attractive to the general reader... is an account of signs which have given warning of a coming earthquake. Underground sounds have been heard, springs have varied in their flow, horses, birds, dogs, and even human beings have been restless for some time before great earthquakes. In his reference to the Riviera earthquake in 1887, Mr. Davison remarks that as premonitions were noted at 130 different places within the central area, "there can be little doubt that they were caused by microseismic movements for the most part insensible to man." In these days of psychical research we think that the author has lost an opportunity for romantic speculation... It is pointed out that the area over which earthquake sounds are heard is variable in different countries. One reason for this is that the limits of audibility vary with different races. From illustrations given it would appear that for certain sounds the Anglo-Saxon ear is more acute than that of the Neapolitan, and very much more than that of the Japanese.

From *Nature* 6 April 1905.

50 YEARS AGO

A new instrument called the 'maser' (microwave amplification by stimulated emission of radiation) has been invented by Prof. C. H. Townes, of the Physics Department, Columbia University, for which the claim is made that it enables time to be measured with an accuracy of one part in 10¹¹. The 'clock' used is an ammonia molecule which radiates an electric dipole spectrum as a set of lines of about 6 mm. wavelength and, as used in the instrument, maintains its frequency to the above order of magnitude. This would be sufficient to enable it to measure variation in the rate of rotation of the earth... the instrument consists of a molecular beam of ammonia molecules which are excited in an electric field and then pass into a tuned cavity-resonator, where they induce each other to radiate by negative absorption... It seems that the method of use is to extract by means of wave-guides the radiation emitted by two 'masers', tuned to different but adjacent frequencies by the ammonia spectrum. When mixed, the beat frequency can then be counted electronically. The 'maser' is also claimed by its inventor to be very effective as an amplifier.

From *Nature* 9 April 1955.

Technology

Hydrogen and hydrates

Ferdi Schüth

It's a potentially explosive issue. How can hydrogen be stored cleanly, efficiently and, above all, safely? One answer would appear to be: take a cage made of water, and add just a little organic solvent.

Finding cheap and easy storage methods will be crucial to establishing hydrogen as a fuel of the future. Clathrates — in which hydrogen molecules are encapsulated or 'occluded' in a cage-like lattice of water molecules — are one option, but until now the pressure required to maintain the stability of such systems has been too high. Lee *et al.*, writing on page 743 of this issue¹, now report that adding small amounts of the common solvent tetrahydrofuran (THF) can substantially reduce this pressure, possibly making a system consisting of hydrogen and water feasible for storage.

Clathrate hydrates are best known for their ability to encapsulate methane, which is assumed to be present in this form in vast amounts on the ocean floor, where the high pressures needed to stabilize the clathrates prevail². In such gas hydrates the water molecules form cages in which 'guest' molecules, such as methane, can be trapped. Some years ago it was discovered³ that hydrogen molecules could also be encapsulated in clathrate structures, albeit at high pressures of about 2 kbar that preclude their practical use for hydrogen storage.

Lee *et al.*¹ describe how this pressure can be

lowered to about 100 bar by 'co-occluding' THF molecules in the clathrate, allowing the storage of up to 4% hydrogen by weight under moderate conditions. Storage capacities of slightly above 1% had previously been shown to be possible with a similar system⁴. Clathrate hydrates are thus comparable to other systems discussed for hydrogen storage, such as light metal hydrides⁵, and are even superior to some, for example cryosystems that use adsorbents with a large surface area (Fig. 1).

The so-called sII clathrate, which is one of the most common gas hydrate structures and is investigated by Lee *et al.*, contains two types of cage — larger ones with a free diameter of about 0.67 nm, and smaller ones of about 0.5 nm (see Fig. 3 in their paper, page 745). For a clathrate to be stable, the large cages must be almost completely filled with suitable guest molecules, while the smaller cages may either remain empty or be filled with guest species. In the pure H₂ clathrate hydrate, four hydrogen molecules are located in the large cages and two in each small cage, giving an overall composition of (2H₂)₂·(4H₂)·17 H₂O, corresponding to a H₂ content of 5% by weight. Florusse *et al.*⁴ showed that the small cavities can be occu-

pied by hydrogen molecules if the clathrate phase is stabilized by THF in the large cages. Two groups — those of Lee in Korea and Ripmeester in Canada — have now found¹ that a stable gas hydrate can be produced by filling only a small fraction of the large cages with THF, provided that the remaining elements conform to the standard pure H₂ clathrate hydrate structure.

Adding just the right amount of THF is a fine balancing act. Above a threshold molar concentration of around 2%, only the small cavities in the clathrate host the hydrogen molecules, and the hydrogen storage capacity remains constant at roughly 2% by weight (Fig. 3 in ref. 1). Below this concentration, hydrogen begins to enter the large cages as well, and the storage capacity increases rapidly to the maximum 4%. At still lower THF concentrations, however, the hydrogen pressure of 120 bar used in the study is no longer sufficient to stabilize the clathrate — in this range the conditions gradually approach the situation in the pure H₂ hydrate, requiring a stabilizing pressure of around 2 kbar.

Analysing hydrogen storage capacities is not simple, as many erroneous results in the literature prove. Here Lee *et al.* have used two independent methods — integration of the signals detected using Raman spectroscopy and volumetric control experiments for selected samples — which are in good agreement with each other. The intensities found in the Raman spectra are independent of the hydrogen content, suggesting that the interaction of the hydrogen molecules with each other and with the water molecules forming the cages is very weak, and that the gas molecules can indeed be considered as occluded.

Could such hydrogen hydrates be developed to a state where they become viable hydrogen-storage systems? The answer is 'maybe'. A storage capacity of 4% falls short of the targets set by the US Department of Energy⁶ and many car manufacturers. Using the clathrates described by Lee *et al.* in hydrogen storage systems for fuel-cell-driven cars would also be hampered by the requirement for a refrigeration system to maintain a temperature of around 0 °C.

However, these clathrates do offer some advantages that may make them attractive for other applications: the storage materials — water and THF — are cheap compared with all other alternatives discussed; the associated environmental and health hazards are relatively low (THF is classified as an 'irritant' in Europe, one of the lower categories with respect to toxicity, and would only be needed in modest amounts); and explosive decomposition is unlikely. Thorough kinetic and thermodynamic studies would be necessary to confirm this last point, but below 0 °C gas hydrates can be metastable, even at ambient pressure (see ref. 7 and references therein).

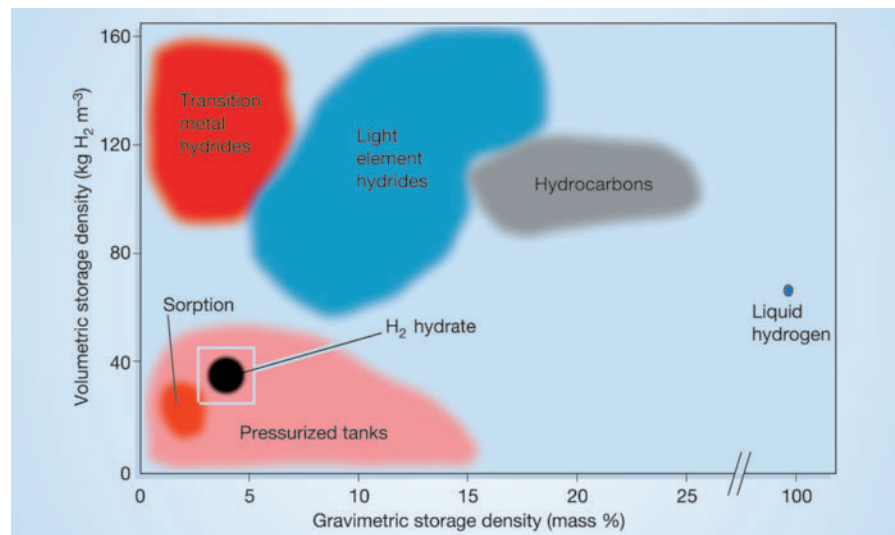


Figure 1 Clathrate storage capacity. Volumetric (mass H₂ per unit volume of storage medium) versus gravimetric (% H₂ storage density) values for different hydrogen storage systems, showing the relative position of the hydrogen hydrate. The density of the H₂ hydrate has not yet been reported but can be estimated at around 0.83 g cm⁻³. This brings the volumetric storage density to about 50% of that of liquid hydrogen. The thermodynamic properties of most of the hydrides in the upper field of the diagram make them unsuitable for reversible hydrogen storage. Hydrocarbons need re-forming and liquid hydrogen needs a refrigeration system (not included in the calculation). Weights and volumes of pressure tanks for pressurized storage are included; actual storage densities depend on tank type. (Diagram simplified from ref. 8.)

A key issue for hydrogen storage materials is that the hydrogenation and dehydrogenation processes should be sufficiently fast. The formation reaction of bulk hydrates is generally slow — owing to diffusion limitations — taking days to weeks. Lee *et al.* have taken a first step towards speeding up this process by dispersing the hydrate on silica beads with a large surface area. This reduces the times for loading and discharge to a matter of hours.

Other hydrate structures are also known, so it may even be possible to discover stable hydrogen hydrates with higher storage capacities or to find more efficient stabilizers than THF. The sII hydrates investigated by Lee *et al.* have among the highest free volumes in known clathrate phases, but the

compositional tuning that they make possible using auxiliary guest species may lead to the discovery of phases with even more attractive properties.

Ferdi Schüth is at the Max-Planck-Institut für Kohlenforschung, 45470 Mülheim an der Ruhr, Germany.

e-mail: schueth@mpi-muelheim.mpg.de

1. Lee, H. *et al.* *Nature* **434**, 743–746 (2005).
2. Sloan, E. D. *Nature* **426**, 353–359 (2003).
3. Mao, W. L. *et al.* *Science* **297**, 2247–2249 (2002).
4. Florusse, L. J. *et al.* *Science* **306**, 469–471 (2004).
5. Schüth, F., Bogdanovic, B. & Felderhoff, M. *Chem. Commun.* 2249–2258 (2004).
6. US Department of Energy Hydrogen Posture Plan www.eere.energy.gov/hydrogenandfuelcells/pdfs/hydrogen_posture_plan.pdf
7. Kuhs, W. F., Genov, G., Staykova, D. K. & Hansen, T. *Phys. Chem. Chem. Phys.* **6**, 4917–4920 (2004).
8. Züttel, A. *Naturwissenschaften* **91**, 157–172 (2004).

Environmental science

Germ theory for ailing corals

Stephen R. Palumbi

Human activities damage coral reef ecosystems. Application of the 'germ theory', proposed more than a century ago for human diseases, could foster action on global environmental ailments such as this.

In 1876, Robert Koch was struggling to convince the world that germs cause disease. Today, environmental degradation is a pervasive planetary disease, but the causes remain shrouded in the same popular murk that made diseases mysterious before the work of Koch and Louis Pasteur. For environmental issues, such as the decline of coral reefs, sceptics demand detailed evi-

dence — we must know the exact cause and show that any proposed cures will work. This is a tall order. Writing in *Science*, however, Pandolfi *et al.*¹ chart the decline of reefs using data from previous work² and some new evidence, and provide a prescription to begin a planet-wide cure.

Some demand that we rigorously prove the cause before acting. Koch faced this

dilemma in the form of scepticism about the germ theory of disease. He responded by deploying the following postulates³ for showing that particular bacteria cause a disease:

- The bacteria must be present in every case of the disease.
- The bacteria must be isolated from the host with the disease and grown in pure culture.
- The specific disease must be reproduced when a pure culture of the bacteria is inoculated into a healthy, susceptible host.
- The bacteria must be recoverable from the experimentally infected host.

These postulates do not always apply, even in medicine³, but they emphasize the critical role of correlational^{1,4} and experimental data^{2,3} in untangling complex problems. These approaches are nearly ubiquitous in the environmental sciences and an environmental version of these rules might be termed 'planetary postulates':

- The cause must be present in every case of the environmental condition.
- The cause must be isolated and known to act by itself to harm the ecosystem.
- The specific condition must be reproduced when the cause is experimentally introduced into the environment.
- The cause must again be verifiably present and active in the affected environment.

Action to address serious environmental issues need not wait until all of these postulates are fulfilled. However, Pandolfi *et al.* meet most of them by showing that reef degradation is a disease of human overuse, as opposed to part of natural cycles. First, fully degraded reefs are always affected by people, because reefs with easier access or longer habitation are more likely to be degraded¹. Second, some human activities harm reefs (Fig. 1) — dynamite fishing, dumping sewage or sediment, dredging corals for cement and unsustainable fishing are well documented⁴. When people are excluded from reefs, fish and invertebrates tend to recover and reef decline tends to reverse⁵. Third, when people are reintroduced to reefs — usually through the collapse of marine reserve enforcement — the condition of over-exploitation returns⁶. Fourth, once people return to reefs, they can be demonstrably shown to be harming them again⁷.

Although the planetary postulates may be met for coral reef damage, human impacts are complex. The culprit may be sedimentation from terrestrial runoff⁸, excessive nutrient input from sewage⁹, the introduction of foreign species or disease-causing organisms¹⁰, overfishing² or global warming¹¹. Analysing any one of these causes in isolation would not survive the strictures of the planetary postulates. But combining all of them, and their ultimate — human — driver seems to fulfil the postulates.

The value of this exercise should be that



Figure 1 Net effect — one manifestation of deleterious human impact on coral reefs. Abandoned nets trap fish and diving birds, and abrade the coral.

diagnosis paves the way for a solution. Pandolfi *et al.* do not know exactly how to cure coral reefs of human overexploitation — no more than Koch, labouring in advance of the discovery of antibiotics, knew how to cure tuberculosis. Instead, they treat the problem as Koch might have done: keep the patients alive by alleviating the symptoms, and reduce exposure to the problem so that they can cure themselves.

To alleviate the symptoms, the most severe impacts on reefs must be reduced. This may mean investment in sewage treatment and abatement of run-off from land; reducing fishing intensity; establishing fully protected reserves; or building buffer zones with limited development near reefs. In the long term it may mean reducing global warming. In the meantime, to keep the patient alive, it is necessary to establish large coral reef parks, such as the Great Barrier Reef Marine Park, as well as small reserves to act as local seed sources.

Finally, the patient needs to heal. Here the sea is ready to help. Marine species have prodigious reproductive abilities — many female fish and invertebrates produce millions of eggs a year. Some corals are virtually immortal and can fragment to produce hundreds of clonal offspring. Movement of tiny larvae can transport species from healthy to damaged reefs. Reefs have recovered from hurricanes, floods, tsunamis and volcanoes. They can also recover from us.

Critics will say that Pandolfi and colleagues' proposed solutions are not socially or technically feasible. Koch faced similar problems. The requirements of his postulates outstripped the abilities of the microbiology and medicine of his day. But Koch did not worry that he could not culture every disease organism, nor that disease germs could not necessarily be killed once identified. Then, as now, progress is attained in steps, and the identification of causes is a key step.

Pandolfi *et al.* make a final plea — don't demand a perfect solution. As an initial goal, it is enough that the health of a reef stops declining and begins to improve. The next step is urgently to develop the science of global ecology to provide a toolbox of more lasting solutions. ■

Stephen R. Palumbi is in the Department of Biological Sciences, Hopkins Marine Station, Stanford University, Pacific Grove, California 93950, USA.

e-mail: spalumbi@stanford.edu

- Pandolfi, J. M. *et al.* *Science* **307**, 1725–1726 (2005).
- Pandolfi, J. M. *et al.* *Science* **301**, 955–958 (2003).
- Harden, V. *Hist. Phil. Life Sci.* **14**, 249–269 (1992).
- Nystroem, M., Folke, C. & Moberg, F. *Trends Ecol. Evol.* **15**, 413–417 (2000).
- Halpern, B. & Warner, R. *Ecol. Lett.* **5**, 361–366 (2002).
- Russ, G. R. & Alcala, A. C. *Coral Reefs* **18**, 307–319 (1999).
- Russ, G. R. & Alcala, A. C. *Naga: ICLARM Quart.* **17**, 8–12 (1994).
- McCulloch, M. *et al.* *Nature* **421**, 727–730 (2003).
- McCook, L. J. *Coral Reefs* **18**, 357–367 (1999).
- Harvell, C. D. *et al.* *Science* **285**, 1505–1510 (1999).
- Knowlton, N. *Proc. Natl Acad. Sci. USA* **98**, 5419–5425 (2001).

Developmental biology

Reproduction in clusters

François Spitz and Denis Duboule

Homeobox genes have some quirky features: they huddle together and tend to be expressed in the order that they appear in their cluster. A new cluster, specific to reproductive development, has now been discovered.

The survival of animal species depends upon the proper development of germ cells: oocytes and sperm. In mammals, this process is tightly coordinated with the development of non-reproductive cells nearby, such as Sertoli cells, which nourish developing sperm cells¹. Changes in the delicate interactions between reproductive and non-reproductive cells are often a source of decreased fertility, so the appropriate regulation in time and space of the underlying genetic determinants must be essential. Writing in *Cell*, MacLean *et al.*² describe how they identified a new set of these genetic determinants — a cluster of homeobox genes that are expressed during the development of germ cells. The clustering of these genes may help to determine their spatial and temporal regulation.

Homeobox genes are so called because they contain a characteristic DNA sequence, the homeobox. They are found in all animals, from mammals to fruitflies, and they are essential for numerous aspects of embryonic development. Until now, however, no homeobox gene cluster had been shown to be directly required for oocyte or sperm maturation, and none was known to be located on the sex chromosomes.

But MacLean and colleagues² have identified a previously unnoticed group of 12 homeobox genes within a small region on the mouse X chromosome. The authors call these genes 'reproductive homeobox X-linked' (RhoX) genes, and show that they define a new homeobox subfamily. Further sequence analyses reveal that the genes fall into three subgroups, which also correspond to their physical alignment along the X chromosome and therefore define three genomic subclusters (α , β and γ ; Fig. 1). The only exception is the *RhoX7* gene, which is located in the β subcluster but whose sequence has more in common with the α subcluster. MacLean *et al.* have also found that the RhoX genes are specifically expressed in male and female reproductive tissues and in the

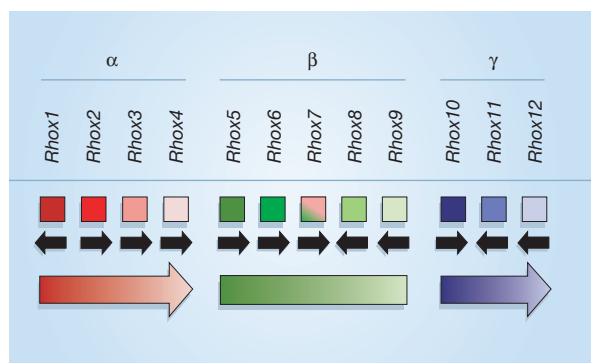


Figure 1 Sperm and gene clusters. MacLean *et al.*² have identified the RhoX gene cluster on the mouse X chromosome. This cluster is organized into three subclusters, α , β and γ . The genes of each subcluster show collinear expression during sperm differentiation — they are expressed in the order that they occur on the chromosome. They are also expressed to different degrees and, in the α and γ subclusters, at different times (graded coloured arrows), such that the first genes in each subcluster are expressed earlier and to a higher maximal level than the next ones. Black arrows denote the direction of gene transcription. The pink shading for the *RhoX7* gene shows that, in sequence terms, it is more closely related to the α than the β subcluster.

placenta. Within the testis, they are mainly expressed in Sertoli cells, some under the control of androgen hormones.

Interestingly, during sperm differentiation, a striking correlation is observed between the genomic organization of the subclusters and the expression of the genes within them. Specifically, as one moves from the start to the end of each subcluster, the genes display a progressive decrease in their efficiency of expression (Fig. 1). In addition, in both the α and the γ subclusters, genes located at the start of the subcluster are activated earlier than genes further along.

These waves of gene activation correlate with changes in the nature of Sertoli cells and their associated germ cells, suggesting that different sets of RhoX genes might control the transition between different stages of maturation. Accordingly, MacLean *et al.* found fewer round and elongated spermatis in mice with mutations in the *RhoX5* gene. So the authors propose that the progressive maturation of sperm is probably controlled by RhoX genes, through their coordinated expression in Sertoli cells.

The regulatory strategy implemented by the RhoX genes is reminiscent of that mediated by other homeobox genes. In mammals, for instance, genes from the Hox family are also clustered and show a temporal sequence

Evolutionary biology

Channels of resistance

Two studies reported in this issue provide striking examples of how biologists are getting to grips with adaptive diversification at the molecular level. They deal with two very different animals — one a marine invertebrate and the other a terrestrial vertebrate.

The softshell clam (*Mya arenaria*) occurs around the Atlantic coast of North America. The clams can become contaminated with saxitoxin, the cause of paralytic shellfish poisoning in humans and economic losses to the shellfish industry. The toxin is produced by 'red tide' algae and finds its way into the clams when the algae are ingested. V. Monica Bricelj *et al.* (*Nature* **434**, 763–767; 2005) show that clams from areas subject to recurrent red tides are relatively resistant to the toxin and tend to accumulate it in their tissues. But clams from unaffected areas have low resistance when exposed to the toxin in the laboratory. These differences were mirrored by the

sensitivity of isolated clam nerve-trunks exposed to the toxin *in vitro*.

To investigate the underlying molecular mechanism, Bricelj *et al.* sequenced the genomic region encoding a putative voltage-gated sodium channel. Such channels sit in cell membranes and regulate ion flow. The authors found a single mutation that correlated with resistance to the toxin, and that results in replacement of a glutamic acid by aspartic acid at a site previously implicated in the binding of saxitoxin. When introduced into a channel from rat brain, this mutation did not affect ion conductance. But the sensitivity of the channel to saxitoxin was greatly reduced owing to a large decrease in the binding affinity of the toxin at the channel pore.

Saxitoxin produced by red-tide algae probably acts as a potent selective agent on the clams, leading to genetic adaptation, the target of selection being genetic variation at a single site in an ion channel.



But this phenomenon is not unique to clams. Saxitoxin is related to another nerve poison called tetrodotoxin (TTX). In some populations of the newt *Taricha granulosa*, individuals accumulate large amounts of TTX in their skin as a defence against garter snakes (*Thamnophis sirtalis*; pictured). As a result, the snakes that prey on toxic newts have evolved high levels of resistance to the toxin. Shana L. Geffeney *et al.* (*Nature* **434**, 759–763;

2005) show that variation in the level of resistance of garter snakes co-evolving with their newt prey can be traced to molecular changes that affect the binding of TTX to — you guessed it — a sodium channel.

So similar mechanisms underlie the adaptation of both softshell clams and garter snakes to regular neurotoxin exposure. Evolution is baroque in its many aspects, but is sometimes more predictable than we imagine.

Rory Howlett

of activation along the head–tail body axis that tightly reflects their order in the cluster³. A related phenomenon is observed for the *tinman*, *bagpipe* and *ladybird late* genes in the 93DE cluster of fruitflies⁴, hinting that 'temporal collinearity' is a widespread mechanism used to coordinate the expression of clustered genes in time. Quantitative collinearity — collinear variations in activation efficiencies, rather than in the temporal order of expression — is also observed for some Hox genes. In the case of RhoX genes, however, the relative activation efficiencies seem to be coordinated only at their peak values, and are not maintained throughout sperm development. So they may merely reflect the genes' time-delayed activation.

The coordinated activation of the RhoX genes also has features in common with the developmental switches described for genes of the β -globin cluster. Embryonic globins are replaced by fetal and then adult forms⁵ in cohorts of maturing blood cells, just as RhoX genes are expressed in a temporal order that coincides with the stages of Sertoli-cell maturation. Thus, both cell types progressively acquire different but related functions as they differentiate.

So how did the RhoX genes evolve? MacLean *et al.* propose² that — like other homeobox genes — they arose through the repeated duplication of an ancestral gene,

and then adopted distinct properties through positive selection. Why, though, have they remained clustered? Duplication is a powerful way to diversify gene functions. But it may also induce some imbalance in gene dosage, which can be detrimental to biological processes that are regulated by delicate concentration equilibria. Keeping duplicated genes in the same genomic region may help to deal with this problem by allowing better coordination of their activation. This could be achieved, for example, if several genes were to compete for the same enhancer regions (DNA sequences that promote gene expression). Enhancers might subsequently evolve into 'switches', selecting the right gene for a given task at a given time. Such sharing of regulatory elements may have constrained the RhoX genes to remain clustered.

However, MacLean *et al.* point out that DNA sequences near the *RhoX5* gene can specifically direct its expression in Sertoli cells, suggesting that some RhoX genes are individually rather than coordinately controlled — at least as far as their cell-type-specific expression goes. The same is true of some Hox genes: when removed from their cluster and placed elsewhere in the genome, they frequently continue to display fairly good spatial expression specificity. But they often lack precise temporal control. So it seems plausible that a strong evolutionary

constraint maintaining both RhoX and Hox gene clustering is the requirement to activate them in a temporal sequence³, as determined by the architecture of these gene clusters. Further research is needed to determine whether androgens (or other hormones) contribute to this temporal activation.

MacLean and colleagues' discovery² of the RhoX genes and their collinear organization and expression provides yet another example of a complex genomic region that somehow coordinates the expression of its genes to regulate a delicate and essential process. It will be interesting to learn more about how the peculiar architecture of the RhoX cluster influences the regulation of these genes and their precise functions, not only throughout germ-cell development, but also in ovaries and the placenta.

François Spitz and Denis Duboule are at the National Research Centre 'Frontiers in Genetics', and the Department of Zoology and Animal Biology, University of Geneva, Sciences III, Quai Ernest Ansermet 30, 1211 Geneva 4, Switzerland. e-mail: Denis.Duboule@zoo.unige.ch

1. Griswold, M. D. *Biol. Reprod.* **52**, 211–216 (1995).
2. MacLean, J. A. II *et al.* *Cell* **120**, 369–382 (2005).
3. Kmita, M. & Duboule, D. *Science* **301**, 331–333 (2003).
4. Jagla, K., Bellard, M. & Frasch, M. *BioEssays* **23**, 125–133 (2001).
5. Stamatoyannopoulos, G. & Grosfeld, F. in *The Molecular Basis of Blood Diseases* (eds Stamatoyannopoulos, G., Majerus, P. W., Perlmutter, R. M. & Varmus, H.) 135–182 (Saunders, Philadelphia, 2001).

Dante's insight into galilean invariance

The poet's vividly imagined flight unwittingly captures a physical law of motion.

In 1632, Galileo described his experience of motion aboard a large ship^{1,2} and exposed in detail the invariance principle, which was then rightly named after him. I suggest that more than three centuries earlier, in *The Divine Comedy*, his fellow countryman Dante Alighieri intuitively grasped what Galileo was later to establish as one of the pillars of modern science.

The Divine Comedy narrates Dante's journey through the hereafter. A remarkable feature of this masterpiece is the vividness of the narration: it is a work of fiction brought to life by the realistic detail. One part deals with Dante's descent from the seventh to the eighth circle, recounted in canto XVII of the *Inferno*. Up to this point, the pilgrim Dante had travelled on foot and occasionally by other means (for example, on Phlegyas' boat, in canto VIII, and on the back of the centaur Nessus, in canto XII).

However, the steepness of the bank between the two circles makes descent possible only by flying—in this case on the back of the monster Geryon. Dante himself reminds us that he is not the first, mythologically speaking, to have flown: Phaeton and Icarus had done it before him. But he is probably one of the first to describe the actual sensation of flying, which he does with an accuracy that is evidence of his extraordinary imagination, as noted by several commentators³.

Sitting on Geryon's back, Dante describes the initial motion of the infernal monster by comparing it to "a little vessel [that] shoves from shore". Geryon then turns and leaves the circle rim, causing Dante to lose his points of visual reference: "I perceived myself / On all sides in the air, and saw extinguished / The sight of everything but of the monster". Then (lines 115–117, written around 1310 and translated by Henry Longfellow in 1867 (ref. 4)): "Onward he goeth, swimming slowly, slowly / Wheels and descends, but I perceive it only / By wind upon my face and from below."

These three lines are a typical example of Dante's ability to summarize several concepts and to describe complex situations in simple statements of a few words. In spite of its apparent simplicity, the tercet reveals some remarkable physical intuition. First, the observer Dante can imagine himself in a frame that a contemporary physicist would define, with fair approximation, as inertial. Although the monster is wheeling, its motion occurs "slowly, slowly" while obeying Virgil's command "Now, Geryon, bestir thyself; / The circles large, and the descent be little". So the flight path corresponds to a wide, slowly sinking spiral and is travelled



Flight of fancy: Gustave Doré's vision of Geryon, who helps Virgil and Dante along their way.

smoothly ("swimming"), or, a physicist would say, almost uniformly.

At the bottom, the monster sets the poet down close to the wall, so the flight path may be assumed to have a diameter similar to that of the cylindrical "abyss". This was estimated by, among others, Galileo⁵ to be 35 Italian miles wide, or about 60 km: this value relies on two precise indications, given in canto XXIX, line 9 and canto XXX, lines 86–87, respectively (which also suggest that Dante was aware of the approximation $\pi \approx 22/7$).

A contemporary physicist could show that, given such dimensions and whatever the speed, the fictitious centrifugal force experienced by the passenger would be much smaller than the surface force due to the apparent wind: no such force is mentioned in the text. Although such a derivation would have been beyond the knowledge of physics in the Middle Ages, Dante did understand how his motion occurred seemingly in a straight line: he indicates its direction, by decomposing the vector describing the apparent wind in terms of its horizontal ("upon my face") and vertical ("from below") components.

Now comes the crucial point. With regard to the motion experienced on the flying monster, in the original tercet Dante says "*ma non me ne accorgo*": he is not aware (or, more accurately, he imagines that he is not aware) of anything but the apparent wind. Dante asserts that, aside from the effect of the wind, his sensation of flying was not dissimilar from being at rest. From a physical point of

view, this 'invariance' is in agreement with the concept expressed by Galileo some three hundred years later. In Galileo's work, the invariance is explicitly related to observations and experimental results. But sitting on the monster's back, Dante could not do much better than rely on and consider his own sensory perceptions.

It is difficult to argue that this descriptive accuracy is accidental. Dante intentionally devised a journey, setting up different scenes and situations to express his message directly or allegorically. His imagined experience of flight is the core of this part of the narration, with the entire scene and landscape being created as a stage for its description. In this instance he seems to avoid allegory, perhaps to allow for a more factual, physical interpretation of the text.

Dante intuitively grasped the concept of invariance but, unlike Galileo, he did not pursue this idea any further. Still, it seems that he was well ahead of his time with regard to the views about the laws of nature held in the Middle Ages.

Leonardo Ricci

Dipartimento di Fisica, Università di Trento,
I-38050 Trento-Povo, Italy
e-mail: ricci@science.unitn.it

- Galilei, G. *Dialogo Sopra i Due Massimi Sistemi del Mondo Tolemaico e Copernicano* (Giulio Einaudi Editore, Torino, 1970).
- Galilei, G. *Dialogue Concerning the Two Chief World Systems, Ptolemaic and Copernican* (transl. Drake, S.; University of California Press, Berkeley, 1953).
- Alighieri, D. *La Divina Commedia* (comm. Sapegno, N.; La Nuova Italia Editrice, Firenze, 1982).
- Alighieri, D. *The Inferno* (transl. Longfellow, H. W.; Barnes & Noble Classics, New York, 2003).
- Galilei, G. *Due lezioni all'Accademia Fiorentina circa la figura, sito e grandezza dell'Inferno di Dante in Le Opere di Galileo Galilei* (Giunti Barbera, Firenze, 1968).

Competing financial interests: declared none.

Anthropology

The earliest toothless hominin skull

The site of Dmanisi in the Eurasian republic of Georgia has yielded striking hominin, faunal and archaeological material as evidence for the presence of early *Homo* outside Africa 1.77 million years ago, documenting an important episode in human evolution. Here we describe a beautifully preserved skull and jawbone from a Dmanisi hominin of this period who had lost all but one tooth several years before death. This specimen not only represents the earliest case of severe masticatory impairment in the hominin fossil record to be discovered so far, but also

Modes of faulting at mid-ocean ridges

W. Roger Buck¹, Luc L. Lavie^{2*} & Alexei N. B. Poliakov³

¹Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York 10964, USA

²Seismological Laboratory, California Institute of Technology, Pasadena, California 91125, USA

³Royal Bank of Canada, 71 Queen Victoria Street, London EC4V 4DE, UK

* Present address: Institute for Geophysics, Jackson School of Geosciences, University of Texas at Austin, 4412 Spicewood Springs Road, #600, Austin, Texas 78759-8500, USA

Abyssal-hill-bounding faults that pervade the oceanic crust are the most common tectonic feature on the surface of the Earth. The recognition that these faults form at plate spreading centres came with the plate tectonic revolution. Recent observations reveal a large range of fault sizes and orientations; numerical models of plate separation, dyke intrusion and faulting require at least two distinct mechanisms of fault formation at ridges to explain these observations. Plate unbending with distance from the top of an axial high reproduces the observed dip directions and offsets of faults formed at fast-spreading centres. Conversely, plate stretching, with differing amounts of constant-rate magmatic dyke intrusion, can explain the great variety of fault offset seen at slow-spreading ridges. Very-large-offset normal faults only form when about half the plate separation at a ridge is accommodated by dyke intrusion.

Faults are quasi-planar weak zones in the lithosphere, the brittle outer shell of the Earth, where earthquakes and tectonic strain are concentrated. To study how faults form it is logical to look at mid-ocean ridges, where faults are constantly forming. It is also particularly important to understand ridge fault systems because they affect major hydrothermal mineral deposits¹ and chemo-synthetic biological communities².

A rift valley flanked by normal faults was the first feature identified as marking the axis of mid-ocean ridges when ridges were discovered 50 years ago³. To explain the 1–2-km-deep, 20–30-km-wide axial valley seen at most slow-spreading ridges (see Fig. 1b, c) was a great challenge of early ridge studies. A consensus emerged that stretching the cold brittle lithosphere at a ridge is what produces a valley^{4,5,6}.

The space generated at the valley by the far-field pull of plate tectonics can be filled by magmatic dyke intrusion at lower stress than is needed for faults to slip and accommodate tension⁷. For this reason, many authors assume that faults form only during periods when no magma is available for dyke intrusion and that the total slip on faults depends on the time interval between dyking events⁸. According to this standard model, all faults at ridges result from tectonic stretching of thin axial lithosphere during amagmatic periods⁹.

In the last decade it has become clear that the stretching model cannot explain the variety of faults seen at ridges. Three specific observations stand out. First, the dip directions of mid-ocean-ridge normal faults show systematic variability as a function of spreading rate. Nearly all faults mapped at slow-spreading centres dip towards the axis, but about half of faults near fast-spreading centres dip towards the axis and the other half dip away⁹. Stretching faults should dip towards the axis.

Second, faults bounding abyssal hills near fast-spreading ridges begin to form ~2–4 km from the axis of normal axial highs¹⁰ and not at the axis, where the seismically imaged lithosphere is thinnest^{11–13}. Lavas cover faults offset very close to the axis¹⁴, but abyssal-hill-bounding faults continue to grow out to 20–30 km from the axis¹⁵.

Third, a completely different type of fault structure was discovered on parts of slow-spreading ridges called ‘oceanic core complexes’ or ‘megamullion’ structures^{16–19} (Fig. 1c). Rather than pillow basalts that are cut by high-angle faults, the core complex surface, sometimes called a detachment, may expose gabbros and mantle rocks that were probably brought up from depths of at least several kilometres. These regions have been likened to

continental metamorphic core complexes where rocks from depths of ~10 km are exposed at the surface. Oceanic core complexes occur mainly at the inside corners of ridge-transform intersections of slow-spreading ridges. The lithosphere of the outside corners looks more like typical slow-spreading lithosphere with moderate-amplitude abyssal hills and nearly continuous exposures of pillow basalts at the surface²⁰. To fit these core complex structures into the standard stretching model, amagmatic spreading with durations of the order of a million years would be needed. Further, the standard model cannot explain the across-ridge asymmetry associated with most oceanic core complexes.

Faults at buoyancy-dominated ridges

Here we develop models of ridge axis deformation in the light of these observations. To deal with faulting at fast-spreading ridges we consider the effect of local buoyancy. Local buoyancy refers to lateral density variation on the scale of the axial region (within ~10 km of the axis). These density variations may be important at fast-spreading ridges because the axial lithosphere is very thin, hot and underlain by partially molten crust^{11,12}. A few kilometres from a fast-spreading axis the lithosphere may be five times thicker than it is on axis and the crust may be solid^{13,21}. The near-off-axis lithospheric thickness of both fast- and slow-spreading ridges may be nearly the same, on the basis of seismic interpretations and thermal models of ridges^{22,23}. However, at slow-spreading ridges there may be much smaller differences between on- and off-axis lithospheric thickness²⁴ (see Fig. 2).

Lateral density variations may relate to the formation of the axial high seen at many fast-spreading centres (Fig. 1a). We adopt the approach of the recently developed accretional curvature model^{25,26}. Unlike other models^{27,28} it does not depend on potentially complex viscous flow under the ridge axis. Also, this model predicts a magnitude and distribution of brittle strain that is consistent with the average observed horizontal fault strain²⁵. Previous numerical treatments of this model²⁶ could not resolve faults and so direct comparisons with observed fault populations have not been possible.

The main simplifying assumption in the model is that fluid magma rises to the level of local isostatic equilibrium at the axis of plate spreading. Local magmatic isostasy at the axis should occur when there is enough magma to accommodate all plate separation. A justification for this assumption comes from seismic results that show a bright reflector at a depth of 0.7–2 km along the axis of most ridges with axial highs^{11,29}. This bright reflector has characteristics

consistent with a magma-filled lens and so is described as an axial magma chamber. Outside this region that is underlain by magma the lithosphere is too strong to respond in a local isostatic manner. The magma and underlying partial melt should accrete to the sides of the adjacent separating plates and get denser as it cools and freezes. This load bends down each plate, and additional accretion forms a plate with a concave upward curvature. As the curved plate passively moves away from the axis it unbends.

To look at fault development near ridges we use a numerical technique that allows for localization of the deformation in a brittle-plastic layer in shear zones simulating faults⁶. Faults are not specified but develop where the stress is sufficient for brittle yielding. The model uses an explicit finite-element approach that efficiently deals with elastic, viscous and brittle-plastic deformation and is set up in a way similar to our previous ridge faulting

studies^{6,30}. Because the buoyancy-dominated (or axial high) case is expected to develop symmetrically, only one side of the ridge is modelled. Magmatic accretion is assumed to account for all plate separation, so no stretching occurs at the axis. Dykes accrete by addition of a column of new elements at the axis every time the off-axis boundary has moved one grid spacing. The top of the new column is set to be at the level of local isostatic equilibrium. To simulate the freezing of partially molten lower crust, elements are also accreted at the non-vertical boundary between the weak axial zone and the brittle lithosphere (Fig. 2a). Deformation occurs because accreted elements are denser than the material in the axial partial melt zone. The density of the axial zone controls the normal stress boundary condition applied to the axial elements and to the level of accretion of the top accreted elements.

For a reasonable density contrast and brittle layer thicknesses the model reproduces both the shape of the axial high and some of the basic characteristics of the flanking faults. As shown in Fig. 3, the fully two-dimensional model fits the general shape of a real axial high. As with previous thin-plate models, the wavelength of the high depends on the off-axis plate thickness and a good fit was found for a 5-km-thick brittle plate. During plate unbending, extensional plastic failure, normal fault nucleation and offset would occur in the top of the plate above a neutral depth—with compressional failure below that depth.

The pattern of faulting in the numerical models of axial-high-related plate unbending is very sensitive to the prescribed rate and amount of weakening with strain. If the rate of strain weakening is low, say 10 MPa of strength loss over a strain of 50%, then no localization into fault-like behaviour is seen. In contrast, for a high rate and large amount of strain weakening, curved faults developed that cut completely through the lithosphere. Bending was concentrated on these widely spaced 'hinge' faults, but did not result in appreciable fault-offset-related surface relief.

Only with limited strength loss that occurred very rapidly with strain does the model develop a fault pattern that bears some resemblance to real axial-high flanking faults. As shown in Fig. 3, a good case had 10 MPa of cohesion loss occurring over 1% of strain. After a phase of fast-strain weakening, continued slow-strain weakening does not affect the bending faults because bending strains are small. Pairs of graben-bounding faults that dipped towards each other developed during plate unbending. The faults did not develop appreciable offset until they were several kilometres away from the axis and they grew steadily out to a

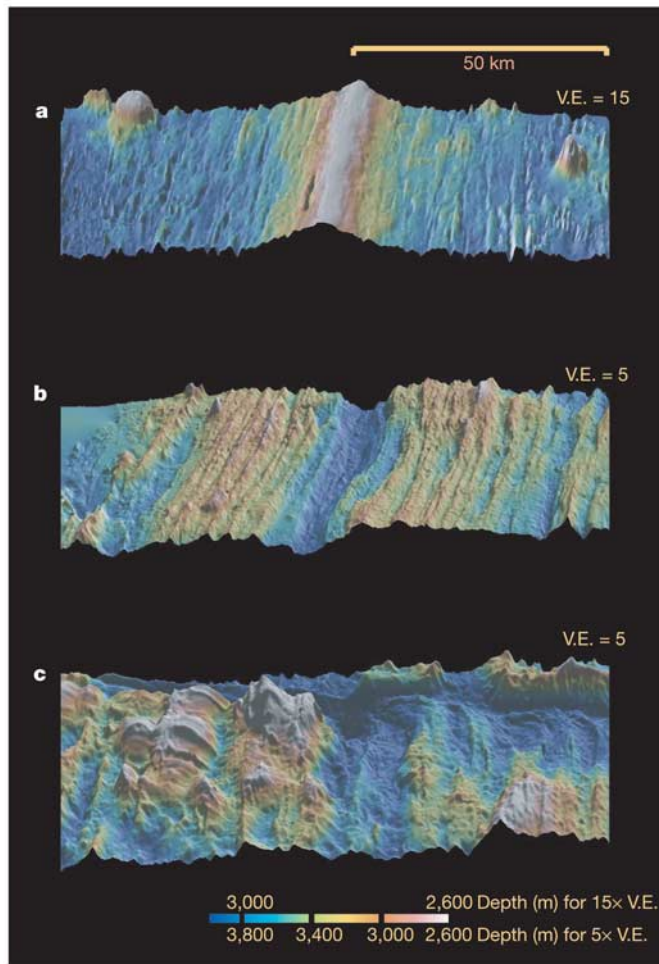


Figure 1 Shaded relief images of bathymetry over three contrasting sections of the mid-ocean ridge system. Each image shows ~40 km along the ridge axis and ~110 km across the axis. The vertical exaggeration (V.E.) is three times greater for the top image than for the other two. **a**, Relief across the axial high at 9° 37' N latitude on the East Pacific Rise showing the small (~50 m) relief of the abyssal hills that parallel the axis. **b**, Much larger abyssal hills are seen parallel to the axial valley along the 115° E longitude segment of the intermediate spreading rate Southeast Indian Ridge. **c**, The intersection of the Mid-Atlantic Ridge, along the centre of image, with the nearly east–west trending Kane transform and fracture zone (on the left and right sides at the top of the image). The image, centred at 23° 25' N latitude, shows that the east side of the ridge has abyssal hills parallel to the axis, as in **b**. Starting ~20 km west of the axis is a ~15-km-wide, 30-km-long structure with great topographic relief and with shorter wavelength 'corrugations' striking perpendicular to the ridge. Profiles orthogonal to these ridge segments, going through the centres of the images, are shown in Figs 3 and 5. Images made with GeoMapApp³⁹ software with multibeam sonar data^{40–42}.

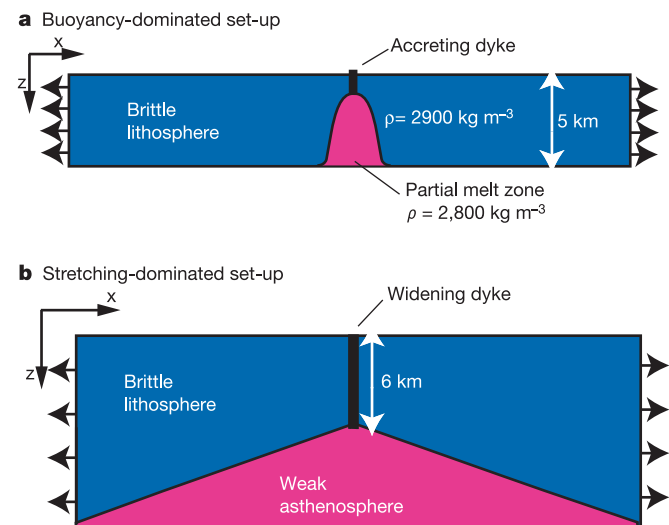


Figure 2 Set-up for numerical models. **a**, Structure used to model a ridge where near-axis buoyancy variations control the topographic relief and faulting. **b**, Set-up for stretching-dominated ridge with a fixed temperature structure.

distance of $\sim 20\text{--}25\text{ km}$. The faults develop over such a wide distance range because that is where progressive unbending occurs, and the faults cease growing where the plates are essentially flattened out. A maximum fault-related relief of $>50\text{ m}$ was developed in the models, in the range of observed values⁹.

Faults at stretching-dominated ridges

To deal with the apparent variety of stretching-related faulting we consider that ridge-axis dyke intrusion at various rates must be considered. Real dykes are intruded over a matter of days with the time between intrusion events measured in years³¹, but in the model we consider the time-averaged rate of dyke intrusion. Magmatic accretion is assumed to occur at a constant rate at all depths through the axial lithosphere. The rate of dyke opening is specified by the fraction of the plate separation rate, M , that is accommodated by magmatic dyke opening. For $M = 0$, dykes account for none of the plate spreading and for $M = 1$, dykes accommodate all the plate separation.

Dykes are assumed to open at one horizontal position in a cross-section of a ridge axis. This assumption is based on the idea of feeding of melt to magma chambers at the centres of ridge segments, for which there is ample evidence^{31,32}. Dykes supplied from a central magma chamber have to open in nearly straight lines to keep the supply of magma flowing along and into an opening dyke. Dyke intrusion may supply much of the heat that keeps the axis hotter and the axial lithosphere thinner than the lithosphere farther from the axis. Solid-state advection below the ridge axis should also affect the thermal structure of the ridge axis, but for simplicity we fix the thermal structure.

A simple geometric argument shows how faults and dykes might interact at a ridge with a fixed position of dyking and fixed thermally defined strength structure. If one fault forms owing to lithospheric stretching then it should initially cut the thinnest axial lithosphere on one side of the axis, shown in Fig. 4. If the fault moves away from the axis into thicker lithosphere, it then becomes harder to continue offsetting that fault even though it is weaker than the surrounding lithosphere. Eventually, it will be easier to form a new fault cutting the axis and the first fault will be abandoned.

Insights from previous models of normal faulting³³ show that when a normal fault forms, one side moves up and away from the fault by about the same amount that the other side moves down and away from the fault. Eventually, the down-dropping side (the hanging-wall block) cannot drop further and stops deforming. The other (footwall) block continues to move up and out as asthenosphere from below accretes to that layer. Because the hanging-wall block no longer deforms, but continues to translate, the active part of the fault moves with the hanging wall. As a result, faults should move with the hanging-wall block, as inferred from these model results. As shown in Fig. 4, the plate to the left of the axis moves with velocity $-V_p/2$, where V_p is the total plate separation rate. The velocity step across the axial dyke is MV_p . Thus, the

hanging-wall block of the fault will move away from the axis at a speed of $V_p(M - 0.5)$.

For $M = 0.5$ the hanging wall of the fault does not move away from the axis, so the fault could build up potentially unlimited offset. The distance L from the axis out to which the fault can remain active should depend on the strength structure of the axis and the fault-weakening parameters. The horizontal component of fault slip rate is $V_p(1 - M)$, so the amount of fault offset before abandonment is $L(1 - M)/(M - 0.5)$, as plotted in Fig. 4. For M close to 1 the maximum fault offset can be small, while for M close to 0.5 the offset should be very large compared to L . This plot is at best a qualitative guide to expected results because L is expected to be a nonlinear function of fault offset^{33,34}.

A numerical model of dyking and stretching

To look at fault development we treat the model of stretching and dyking using the same numerical technique used for the buoyancy cases, except that we allow for asymmetric deformation across the axis. Dyke opening is simulated with a vertical column of special elasto-plastic dyke elements that are made to widen at a constant rate. The dyke column is placed at the axis of the assumed symmetric ridge thermal and strength structure (Fig. 2b). The top of the dyke is set at the height of the adjacent elements, which are free to move up or down in the course of a calculation. The base of the dyke column is placed within the weak asthenosphere, which always deforms via viscous creep (see Supplementary Information for details).

To approximate the strength structure of slow-spreading ridges we assumed uniform elastic coefficients, pressure- and strain-dependent Mohr Coulomb brittle-plastic failure stresses and temperature-dependent non-newtonian viscosity (Supplementary Information).

Several dozen numerical cases were run with various values of M , strain weakening rates and numerical grid sizes. The results of two cases with the same strength structure, grid size and strain-weakening parameters are shown in Fig. 5. These cases show that the normal fault offset varies greatly as a function of M and that for $M = 0.5$ we can get a very large fault offset, as predicted by our kinematic conceptual model. For $M = 0.95$, the model also generates a fairly symmetric pattern of mainly inward-dipping, small-offset faults, and a symmetric axial valley.

For $M = 0.5$ the model produced two large faults with offsets 20 to 30 km on one side of the spreading axis and a series of small faults on the other. On the opposite side of the axis, a series of small-offset normal faults develop, along which the lithosphere moves up from the deep axial valley region. We believe that these small faults, which were not part of our kinematic conceptual model, accommodate a small part of the tectonic stretching across the axis. This small stretching on one side of the axis will contribute to migration of the large-offset faults away from the axis on the other side of the axis. This may account for the limited offset of the large faults seen in this case.

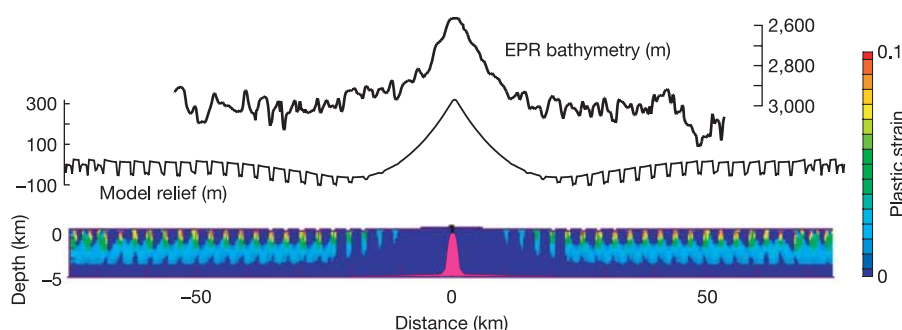


Figure 3 Results for buoyancy-dominated ridge models. The shape of the model axial high and the spacing and relief of the faults is similar to that observed at $9^\circ 37' \text{ N}$ on the

East Pacific Rise (from Fig. 1a). Half the bending faults dip towards the axis and half away, which is also consistent with observations.

Structures developed in association with model large-offset faults are similar to the oceanic core complexes seen at more than twenty ridge-transform intersections on slow-spreading ridges^{17–19}. Besides having a similar range of relief for the axial valley and the inside-corner high (2 km), the model topography has other characteristics observed on topographic profiles for an ocean core complex (Fig. 5b): (1) a domed shape with a flat-lying abandoned-fault footwall; and (2) the characteristic shape of the fault breakaway. Also, for $M \sim 0.5$ nearly all the magmatic accretion occurs on the side of the axis with small-offset faults, as may happen at outside corners of some ridge-transform intersections. Finally, although it is not well resolved in our numerical models, the bending associated with large fault offsets should result in small-offset, high-angle faults. Such distributed small-offset, high-angle normal faults are seen on the surface of oceanic core complexes^{18,19}.

To get the large-offset faults that may characterize oceanic core complexes our models require that three conditions be met. First, M has to be close to 0.5. Next, the fault weakening must be large and

more than half of the weakening has to occur moderately slowly with strain. Finally, the lithosphere cannot thicken very fast with distance from the axis.

Magma supply and fault mode

A single set of fault strain-weakening parameters reproduced the essential features of both stretching- and bending-related faults. Two phases of strain weakening were needed. Nearly instantaneous strength reduction had to occur, once the yield stress was reached, for bending to produce small-offset, inward- and outward-dipping faults similar to those seen near axial highs. Slower strain weakening had to occur to get larger-offset faults seen at some slow-spreading centres. For both the buoyancy- and stretching-dominated model cases the same strain-weakening parameters were used: with fast weakening for the first percent of strain and slow strain weakening for greater strain. The fast strain weakening may correspond to the sudden loss of cohesion as a fault break forms. The continued weakening may relate to wearing and widening of the fault zone with greater offset. It is possible that these strain-weakening parameters may also describe how faults form in other tectonic environments, such as strike-slip and thrust-faulted regions.

The numerical models developed here show how distant loads and local loads may produce very different modes of normal faulting at mid-ocean ridges. All loads driving plate motion and deformation are produced by lateral density variations in the Earth. Stretching is related to loads and motions at the scale of plate tectonics. The tectonic pull caused by cooling and contraction of plates that float on a weak asthenosphere can contribute to stretching by transmitting stresses over long distances. Local loads, related to magmatic accretion and near-axis cooling, may drive

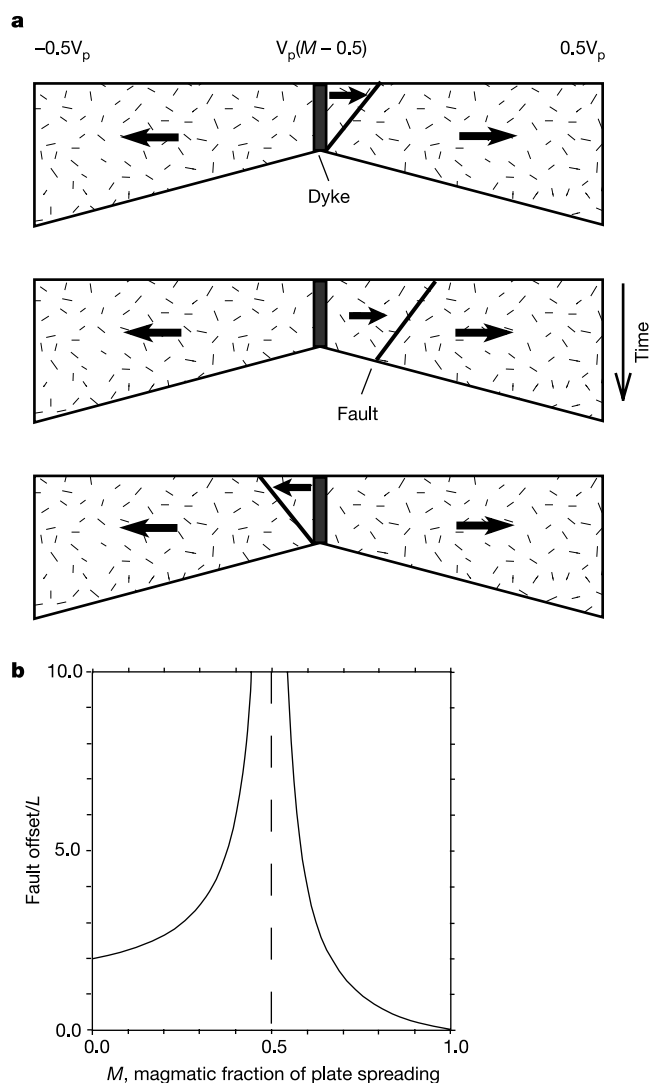


Figure 4 Illustration of kinematic stretching fault model. **a**, Cartoon showing how the hanging-wall block of a fault may migrate during ridge stretching as a function of the fraction, M , of plate separation accommodated by magmatic dyke opening. The velocity of the hanging-wall block relative to the axis is $V_p(M - 0.5)$, where V_p is the full spreading rate. The position of the hanging-wall block should control the fault location. For $M > 0.5$ the hanging wall moves away from the axis, but for $M = 0.5$ the hanging wall does not move laterally with time. **b**, Plot of estimated fault offset as a function of M for the kinematic stretching and dyking ridge model. L is the maximum off-axis distance of fault activity.

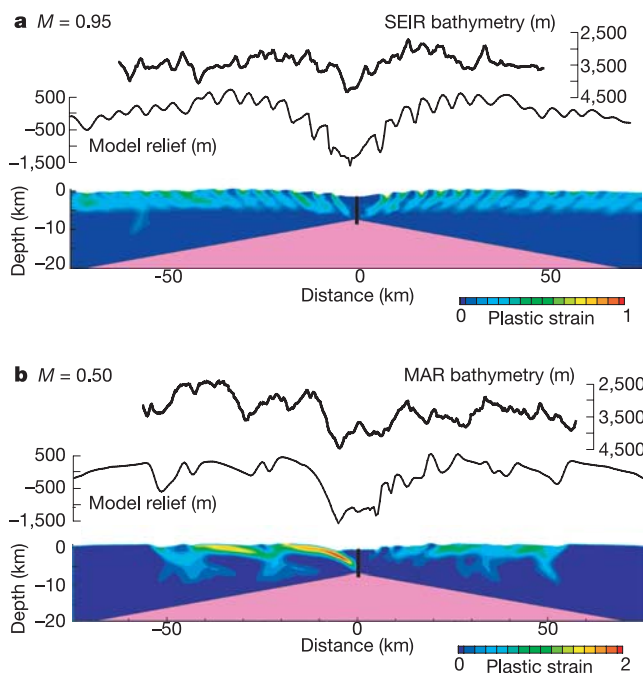


Figure 5 Results for stretching-dominated ridge models. Two model cases are shown in **a** and **b** for different values of M , but with the same lithospheric structure. **a**, Note the fairly symmetric axial valley and the moderate fault offsets for $M = 0.95$ which matches the sense of the bathymetric profile from the 115° E segment of the Southeast Indian Ridge (from Fig. 1b). Reducing the model lithospheric thickness by about one-third produces a better fit to this bathymetry. **b**, Faults with very large offset occur for $M = 0.5$. The model topographic profile is similar to that observed near the intersection of the Mid-Atlantic Ridge and the Kane transform (from Fig. 1c). The model is consistent with the large offset faults seen at the inside corners of such slow-spreading segments, as well as with the asymmetry in magmatic accretion, because most magmatic accretion takes place on the side of the ridge with smaller fault offsets.

lithospheric bending and faulting at any ridge. Bending-related faulting is the dominant faulting mode at fast- and perhaps at some intermediate-spreading centres.

The pattern of faulting produced by stretching a spreading centre is controlled by the rate of magmatic accretion. Temporal variations in magma supply on a very long timescale are not required to produce large-offset faults. The across-axis strength structure can also affect the pattern of faulting, although the models presented here do not deal with different strength structures. For M less than 1, the thermal advection related to fault offset may affect the ways faults continue to slip, but this is a complex question related to hydrothermal cooling of lithosphere³⁵.

We have considered only end-member cases to demonstrate these ideas, so we can only speculate about the behaviours of ridge segments with a range of spreading rates and magma supply. Many intermediate-spreading ridge segments fit reasonably well into our end-member categories, some being buoyancy-dominated and some stretching-dominated. At the slowest-spreading ridges, sometimes called ultraslow-spreading ridges³⁶, an extreme range of along-axis variations in magma abundance is seen on a segment scale (~ 100 km)^{36,37,38}. Unlike typical spreading segments, magma-starved parts of ultraslow-spreading segments can be oblique to the spreading direction. The obliquity of nearly amagmatic parts of ultraslow-spreading ridges may be directly tied to the lack of dyke intrusions coming from a magma chamber at the centre of a segment.

In these models, even small-magnitude stretching (that is M just less than 1) produces large axial valley relief. Maximum relief increases with increasing model axial lithospheric thickness, on the basis of models not presented here. Increased axial valley relief seen at slow-spreading segment ends compared to segment centres correlates with thicker lithosphere at the ends than at centres²⁴.

If our view of the kinematics of the hanging-wall block bounded by a fault and an axial dyke (Fig. 4) for $M \approx 0.5$ is correct, then it suggests a simple explanation for oceanic core complex formation on the inside corner of a ridge-transform intersection. Having this block on the inside of the ridge-transform intersection minimizes the shear strain across the transform. An outside-corner core complex would require slip on the extension of the transform outside the ridge-transform intersection.

The great variations in the fault pattern along many 'normal' slow-spreading segments (for example, Fig. 1) appear to require very different rates of magmatic dyke intrusion along those segments. There may be a central, magma-rich part of these segments where the magma supply is nearly enough to accommodate plate separation (M just below 1) with an adjacent magma-poor part of the segment where $M \approx 0.5$. Our work shows that differences in magmatic input to dykes may produce the observed differences in faults. Understanding why the magmatic accommodation of plate separation varies so much along ridge segments, and why it seems to occur in various modes rather than vary smoothly, is a clear challenge to understanding how spreading centres work. □

Received 8 September 2004; accepted 12 January 2005; doi:10.1038/nature03358.

- Hannington, M. D., Jonasson, I. R., Herzig, P. M. & Petersen, S. in *Seafloor Hydrothermal Systems: Physical, Chemical, Biological, and Geological Interactions* (eds Humphris, S. E., Zierenberg, R. A., Mullineaux, L. S. & Thomson, R. E.) 115–157 (American Geophysical Union, Washington DC, 1995).
- Van Dover, C. L. Evolution and biogeography of deep-sea vent and seep invertebrates. *Science* **295**, 1253–1257 (2002).
- Heezen, B. C. The rift in the ocean floor. *Sci. Am.* **203**, 99–106 (1960).
- Tapponnier, P. & Francheteau, J. Necking of the lithosphere and the mechanics of slowly accreting plate boundaries. *J. Geophys. Res.* **83**, 3955–3970 (1978).
- Lin, J. & Parmentier, E. M. A finite amplitude necking model of rifting in brittle lithosphere. *J. Geophys. Res.* **95**, 4909–4923 (1990).
- Poliakov, A. N. B. & Buck, W. R. in *Faulting and Magmatism at Mid-Ocean Ridges* (eds Buck, W. R., Delaney, P. T., Karson, J. A. & Lagabriele, Y.) 305–324 (American Geophysical Union, Washington DC, 1998).
- Price, N. J. & Cosgrove, J. W. *Analysis of Geological Structures* 1–452 (Cambridge Univ. Press, Cambridge, UK, 1990).
- Thatcher, W. & Hill, D. P. A simple model for fault generated morphology of slow-spreading mid-oceanic ridges. *J. Geophys. Res.* **100**, 561–570 (1995).

- Carbotte, S. M. & Macdonald, K. C. Causes of variation in fault-facing direction on the ocean floor. *Geology* **18**, 749–752 (1990).
- Macdonald, K. C., Fox, P. J., Alexander, R. T., Pockalny, R. & Gente, P. Volcanic growth faults and the origin of the Pacific abyssal hills. *Nature* **380**, 125–129 (1996).
- Detrick, R. S. *et al.* Multichannel seismic imaging of the crustal magma chamber along the East Pacific Rise. *Nature* **326**, 35–41 (1987).
- Vera, E. E. *et al.* The structure of 0- to 0.2-m.y.-old oceanic crust at 9°N on the East Pacific Rise from expanded spread profiles. *J. Geophys. Res.* **95**, 15529–15556 (1990).
- Dunn, R. A. *et al.* Three-dimensional seismic structure and physical properties of the crust and shallow mantle beneath the East Pacific Rise at 9°N. *J. Geophys. Res.* **105**, 23537–23556 (2000).
- Karson, J. A. *et al.* Structure of the uppermost fast-spread oceanic crust exposed at the Hess deep: Implications for subaxial processes at the East Pacific Rise. *Geochem. Geophys. Geosyst.* **2**, 1002, doi:10.1029/2001GC000155 (2001).
- Alexander, R. T. & Macdonald, K. C. Sea Beam, Sea MARC II and ALVIN based studies of faulting on the East Pacific Rise 9°20' N–9°50' N. *Mar. Geophys. Res.* **18**, 557–587 (1996).
- Karson, J. A. *et al.* Along-axis variations in seafloor spreading in the MARK area. *Nature* **328**, 681–685 (1987).
- Cann, J. R. *et al.* Corrugated slip surfaces formed at ridge-transform intersections on the Mid-Atlantic Ridge. *Nature* **385**, 329–332 (1997).
- Tucholke, B. E., Lin, J. & Kleinrock, M. Megamullions and mullion structure defining oceanic metamorphic core complexes on the Mid-Atlantic Ridge. *J. Geophys. Res.* **103**, 9857–9866 (1998).
- Blackman, D. K. *et al.* Geology of the Atlantis Massif (MAR 30°N): Implications for the evolution of an ultramafic core complex. *Mar. Geophys. Res.* **23**, 443–469 (2004).
- Tucholke, B. E. & Lin, J. A geologic model for the structure of ridge segments in slow-spreading ocean crust. *J. Geophys. Res.* **99**, 11937–11958 (1994).
- Crawford, W. C., Webb, S. C. & Hildebrand, J. A. Estimation of shear velocities in the oceanic crust from compliance measurements by two-dimensional finite difference modeling. *J. Geophys. Res.* **103**, 9895–9916 (1998).
- Purdy, G. M., Kong, L. S. L., Christeson, G. L. & Solomon, S. C. Relationship between spreading rate and the seismic structure of mid-ocean ridges. *Nature* **355**, 815–817 (1992).
- Phipps Morgan, J. & Chen, Y. J. The genesis of oceanic crust: magma injection, hydrothermal circulation and crustal flow. *J. Geophys. Res.* **98**, 6283–6297 (1993).
- Barclay, A. H., Toomey, D. R. & Solomon, S. C. Microearthquake characteristics and crustal V_p/V_s structure at the Mid-Atlantic Ridge, 35°N. *J. Geophys. Res.* **106**, 2017–2034 (2001).
- Buck, W. R. Accretional curvature of lithosphere at magmatic spreading centers and the flexural support of axial highs. *J. Geophys. Res.* **106**, 3953–3960 (2001).
- Shah, A. & Buck, W. R. Causes for axial high topography at mid-ocean ridges and the role of crustal thermal structure. *J. Geophys. Res.* **106**, 30865–30880 (2001).
- Kuo, B. Y., Forsyth, D. W. & Parmentier, E. M. Flexure and thickening of the lithosphere at the East Pacific Rise. *Geophys. Res. Lett.* **13**, 681–684 (1986).
- Eberle, M. A. & Forsyth, D. W. An alternative, dynamic model of the axial topographic high at fast spreading ridges. *J. Geophys. Res.* **103**, 12309–12320 (1998).
- Hooff, E. E., Detrick, R. S. & Kent, G. M. Seismic structure and indicators of magma budget along the Southern East Pacific Rise. *J. Geophys. Res.* **102**, 27319–27340 (1997).
- Buck, W. R. & Poliakov, A. N. B. Abyssal hills formed by stretching oceanic lithosphere. *Nature* **392**, 272–275 (1998).
- Einarsson, P. & Brandsdóttir, B. Seismological evidence for lateral magma intrusion during the July 1978 deflation of the Krafla volcano in NE-Iceland. *J. Geophys. Res.* **87**, 160–165 (1980).
- Magde, L. & Sparks, D. W. Three dimensional mantle upwelling, melt generation and melt migration beneath segmented slow spreading ridges. *J. Geophys. Res.* **102**, 20571–20583 (1997).
- Lavier, L. L., Buck, W. R. & Poliakov, A. N. B. Factors controlling normal fault offset in an ideal brittle layer. *J. Geophys. Res.* **105**, 23431–23442 (2000).
- Buck, W. R. Effect of lithospheric thickness on the formation of high- and low-angle normal faults. *Geology* **21**, 933–936 (1993).
- Lavier, L. L. & Buck, W. R. Half graben versus large-offset low-angle normal fault: The importance of keeping cool during normal faulting. *J. Geophys. Res.* **107**, 10.1029/2001JB000513 (2002).
- Dick, H. J. B., Lin, J. & Schouten, H. An ultraslow-spreading class of ocean ridge. *Nature* **426**, 405–412 (2003).
- Cochran, J. R., Kurras, G. H., Edwards, M. H. & Coakley, B. J. The Gakkel Ridge: Bathymetry, gravity anomalies and crustal accretion at extremely slow spreading rates. *J. Geophys. Res.* **108**, doi:10.1029/2002JB001830 (2003).
- Michael, P. J. *et al.* Magmatic and amagmatic seafloor generation at the ultraslow-spreading Gakkel Ridge, Arctic Ocean. *Nature* **423**, 956–961 (2003).
- Haxby, W. *GeoMapApp version 1.2_02*; (<http://www.GeoMapApp.org/>) (Marine Geosciences Data Management System, downloaded August 2004).
- Cochran, J. R. *et al.* The Southeast Indian Ridge between 88°E and 120°E: Gravity anomalies and crustal accretion at intermediate spreading rates. *J. Geophys. Res.* **102**, 15463–15487 (1997).
- Macdonald, K. C. *et al.* The East Pacific Rise and its flanks, 8–18°N: History of segmentation, propagation and spreading direction based on SeaMARC II and SeaBeam studies. *Mar. Geophys. Res.* **14**, 299–344 (1992).
- Gente, P. *et al.* Characteristics and evolution of the segmentation of the Mid-Atlantic Ridge between 20 degrees N and 24 degrees N during the last 10 million years. *Earth Planet. Sci. Lett.* **129**, 55–71 (1995).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements Work supported by the National Science Foundation. We thank W. Haxby for help with images and J. Karson for comments on the text.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.R.B. (buck@ldeo.columbia.edu).

Generation and annotation of the DNA sequences of human chromosomes 2 and 4

LaDeana W. Hillier¹, Tina A. Graves¹, Robert S. Fulton¹, Lucinda A. Fulton¹, Kymberlie H. Pepin¹, Patrick Minx¹, Caryn Wagner-McPherson¹, Dan Layman¹, Kristine Wylie¹, Mandeep Sekhon¹, Michael C. Becker¹, Ginger A. Fewell¹, Kimberly D. Delehaunty¹, Tracie L. Miner¹, William E. Nash¹, Colin Kremitzki¹, Lachlan Oddy¹, Hui Du¹, Hui Sun¹, Holland Bradshaw-Cordum¹, Johar Ali¹, Jason Carter¹, Matt Cordes¹, Anthony Harris¹, Amber Isak¹, Andrew van Brunt¹, Christine Nguyen¹, Feiyu Du¹, Laura Courtney¹, Joelle Kalicki¹, Philip Ozersky¹, Scott Abbott¹, Jon Armstrong¹, Edward A. Belter¹, Lauren Caruso¹, Maria Cedroni¹, Marc Cotton¹, Teresa Davidson¹, Anu Desai¹, Glendoria Elliott¹, Thomas Erb¹, Catrina Fronick¹, Tony Gaike¹, William Haakenson¹, Krista Haglund¹, Andrea Holmes¹, Richard Harkins¹, Kyung Kim¹, Scott S. Kruchowski¹, Cynthia Madsen Strong¹, Neenu Grewal¹, Ernest Goyea¹, Shunfang Hou¹, Andrew Levy¹, Scott Martinka¹, Kelly Mead¹, Michael D. McLellan¹, Rick Meyer¹, Jennifer Randall-Maher¹, Chad Tomlinson¹, Sara Dauphin-Kohlberg¹, Amy Kozlowski-Reilly¹, Neha Shah¹, Sharhonda Swearengen-Shahid¹, Jacqueline Snider¹, Joseph T. Strong¹, Johanna Thompson¹, Martin Yoakum¹, Shawn Leonard¹, Charlene Pearman¹, Lee Trani¹, Maxim Radionenko¹, Jason E. Waligorski¹, Chunyan Wang¹, Susan M. Rock¹, Aye-Mon Tin-Wollam¹, Rachel Maupin¹, Phil Latreille¹, Michael C. Wendt¹, Shiao-Pyng Yang¹, Craig Pohl¹, John W. Wallis¹, John Spieth¹, Tamberlyn A. Bieri¹, Nicolas Berkowicz¹, Joanne O. Nelson¹, John Osborne¹, Li Ding¹, Rekha Meyer¹, Aniko Sabo¹, Yoram Shotland¹, Prashant Sinha¹, Patricia E. Wohldmann¹, Lisa L. Cook¹, Matthew T. Hickenbotham¹, James Eldred¹, Donald Williams¹, Thomas A. Jones², Xinwei She³, Francesca D. Ciccarelli⁴, Elisa Izaurralde⁴, James Taylor⁵, Jeremy Schmutz⁶, Richard M. Myers⁶, David R. Cox^{6*}, Xiaoqiu Huang⁷, John D. McPherson^{1*}, Elaine R. Mardis¹, Sandra W. Clifton¹, Wesley C. Warren¹, Asif T. Chinwalla¹, Sean R. Eddy², Marco A. Marra^{1*}, Ivan Ovcharenko⁸, Terrence S. Furey⁹, Webb Miller⁵, Evan E. Eichler³, Peer Bork⁴, Mikita Suyama⁴, David Torrents⁴, Robert H. Waterston^{1*} & Richard K. Wilson¹

¹Genome Sequencing Center, Washington University School of Medicine, Campus Box 8501, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

²Howard Hughes Medical Institute and Department of Genetics, Washington University School of Medicine, St. Louis, Missouri 63110, USA

³Genome Sciences, University of Washington School of Medicine, Seattle, Washington 98195, USA

⁴EMBL, Meyerhofstrasse 1, Heidelberg 69117, Germany

⁵Center for Comparative Genomics and Bioinformatics, Departments of Biology and Computer Science, Pennsylvania State University, University Park, Pennsylvania 16802, USA

⁶Stanford Human Genome Center, Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA

⁷Department of Computer Science, Iowa State University, Ames, Iowa 50011-1040, USA

⁸EEBI Division and Genome Biology Division, Lawrence Livermore National Laboratory, Livermore, California 94550, USA

⁹Center for Biomolecular Science and Engineering, University of California, Santa Cruz, California 95064, USA

* Present addresses: Perlegen Sciences Inc., 2021 Stierlin Court, Mountain View, California 94043, USA (D.R.C.); Baylor College of Medicine, 1 Baylor Plaza, Human Genome Sequencing Center, N1519, Houston, Texas 77030, USA (J.D.M.); Genome Sciences Centre, British Columbia Cancer Agency, 600 West 10th Avenue, Room 3427, Vancouver, British Columbia V5Z 4E6, Canada (M.A.M.); Department of Genome Sciences, Box 357730, University of Washington, 1705 NE Pacific Street, Seattle, Washington 98195-7730, USA (R.H.W.).

Human chromosome 2 is unique to the human lineage in being the product of a head-to-head fusion of two intermediate-sized ancestral chromosomes. Chromosome 4 has received attention primarily related to the search for the Huntington's disease gene, but also for genes associated with Wolf-Hirschhorn syndrome, polycystic kidney disease and a form of muscular dystrophy. Here we present approximately 237 million base pairs of sequence for chromosome 2, and 186 million base pairs for chromosome 4, representing more than 99.6% of their euchromatic sequences. Our initial analyses have identified 1,346 protein-coding genes and 1,239 pseudogenes on chromosome 2, and 796 protein-coding genes and 778 pseudogenes on chromosome 4. Extensive analyses confirm the underlying construction of the sequence, and expand our understanding of the structure and evolution of mammalian chromosomes, including gene deserts, segmental duplications and highly variant regions.

Less than 50 years after the human diploid number was established, the reference human genome sequence was announced¹. Detailed accounts of the sequences of individual chromosomes are now providing great insights into genomic structure and evolution. Here we present our analysis of the sequence of human chromosomes 2 and 4. For chromosome 2, we analyse the region containing the ancestral chromosome fusion event² and describe possible mechanisms for the inactivation of the vestigial centromere. For chromosome 4, we discover some regions with the lowest and highest (G+C) content in the human genome, as well as the putative largest 'gene deserts'. Analyses of highly variant regions found on these chromosomes have also allowed us to investigate their origins.

Generation of the chromosome sequences

Chromosomes 2 and 4 were sequenced using a clone-by-clone shotgun sequencing strategy³ supported by the bacterial artificial chromosome (BAC)-based whole genome physical map⁴. The quality of the sequences was determined to exceed the 99.99% accuracy standard^{1,5,6}. On both chromosomes, sequences extend into the centromere and reach the p-arm telomere⁷ (Supplementary Table 1). Attempts were made to close all remaining gaps⁸ (Supplementary Methods), with an emphasis on using newly available fosmid libraries. Seventy-six clones were selected from the fosmid end placements¹, which added ~500 kb of sequence (included in release hg17/build35). On the basis of size estimates of remaining

gaps, the available sequence represents more than 99.6% of the total euchromatic sequence.

The integrity of the underlying clone sequences and their assembly into chromosomes were verified by comparisons of an *in silico* digest of each finished sequence to the restriction digests of the clone DNA, and comparison of the full assembly sequence against the underlying fingerprint data. In this way, we directly confirmed more than 99.99% of the testable bands. In addition, by examining the placements of BAC, fosmid and plasmid⁹ paired end sequences, and by comparing the order of BAC end placements to the order of the BACs within the fingerprint map⁴, we confirmed the overall consistency of the map, sequence and assembly (Supplementary Methods).

Comparison to physical and genetic maps

We have used the Genethon¹⁰, Marshfield¹¹ and deCODE¹² micro-satellite-based maps, GeneMap99¹³ and TNG¹⁴ radiation hybrid maps, the Whitehead yeast artificial chromosome (YAC)-based map and the BAC physical map⁴ to evaluate the completeness and order of assembled sequences for chromosomes 2 and 4. Only a small number (<1%) of sequence-tagged sites (STSs) were not identified in the existing sequence; these included STSs from repetitive sequences, sequence polymorphisms and regions within known sequence gaps.

To evaluate the assembly of the sequence, the chromosomal sequence positions of the deCODE STSs were plotted relative to their established map positions (Supplementary Fig. 1). When examining local ordering of the deCODE marker set (the average spacing of which is ~1 per 600 kb), there were no long-range disagreements, and a local inversion of marker pairs was found in only one of the placements. Taken together, these data provide strong support for the presented sequences of chromosomes 2 and 4.

General features

We analysed the chromosome 2 and 4 sequences for interspersed repeat content (Supplementary Table 2), (G+C) content, recombination and the presence of CpG islands. For both chromosomes, the short interspersed nucleotide element (SINE) content is lower than the autosomal average. The long interspersed nucleotide element (LINE) content is higher than average for both chromosomes, with chromosome 4 containing the highest percentage across all autosomes.

The (G+C) content of chromosomes 2 (40.2%) and 4 (38.2%) is lower than that of the genome as a whole (41%). On chromosome 2, the lowest (G+C) content (29.9%) is found in a 40-kb region containing no known genes. Over 19% of chromosome 4 has a (G+C) content of less than 35%, compared with 9% in the genome as a whole. However, one of the highest (G+C) content windows in the genome (72.7%) is also found on chromosome 4.

We identified 1,662 CpG islands in chromosome 2 (7 per Mb), and 1,004 CpG islands in chromosome 4 (5.4 per Mb), each with an average length of ~800 bp. Chromosome 4 has the lowest density of predicted CpG islands of any human chromosome. For the gene sets presented below, 67% (chromosome 2) and 64% (chromosome 4) had overlapping CpG islands, consistent with other estimates^{15,16}.

Chromosomes 2 and 4 have the lowest average recombination rate of any of the chromosomes (1.09 compared with the genome average of 1.31). Each chromosome contains large regions with little or no recombination (Supplementary Table 3). For chromosome 2, these regions include the candidate tumour suppressor gene *LRP1B*¹⁷, which spans 500 kb, the *ZAP70* gene¹⁸ and *TTN*, the longest coding sequence in the human genome, spanning 280 kb and encoding a 2,993 kDa protein. On chromosome 4, a recombination cold spot was identified at ~42 Mb, the location of the paired-like homeobox 2b gene (*PHOX2B*). As expected, the primary hot spots are located near the telomeres, where recombination rates are five- to tenfold higher than average.

Protein-coding genes: known genes

Creation of protein-coding gene indices for chromosomes 2 and 4 exploited both the increasing number of available human messenger RNAs and the extensive orthology with mouse (defined for 94% of chromosome 2 and 97% of chromosome 4), in a hierarchical approach similar to that used for human chromosome 7 (ref. 8). First, a collection of 1,448 (chromosome 2) and 820 (chromosome 4) human mRNAs from RefSeq¹⁹ and the Mammalian Gene Collection (MGC)²⁰ were assigned to the sequence and manually curated, resulting in 1,032 non-overlapping mRNAs on chromosome 2 and 633 on chromosome 4. On the basis of alignments with known mRNAs, 25% of the known genes on these chromosomes have alternative splice forms. However, we estimate that this percentage rises to as high as 85% on the basis of gene structure prediction using expressed-sequence-tag (EST) data and all vertebrate mRNAs with an average number of 5 transcripts per gene. Of the known genes, 89% of those on chromosome 2 and 86% of those on chromosome 4 had an associated poly(A) signal. All known mRNAs previously mapped to these chromosomes were successfully identified, confirming the completeness of the available sequence.

Alignment of the human mRNA set (RefSeq and MGC) against the genome revealed both insights and potentially confounding artefacts. For example, detailed examination identified a set of 36 genes (Supplementary Table 4) predicted from the known mRNAs that had no match at all to the mouse or rat genomes or their protein sets; 26 of these had no similarity to the non-redundant protein database. Sixteen of the 26 genes were single-exon genes, and might be fortuitous open reading frames (ORFs) in untranslated regions (UTRs). Eleven of these single-exon genes had ORFs of more than 300 bp (the average across the 16 single-exon genes was 350 bp). The remaining ten were multi-exon genes. To test whether these were novel human genes, we obtained sequences from other primates for six of the ten multi-exon genes. For five out of six genes, there was an ORF through the re-sequenced exon in each of the equivalent primate genes. In the sixth gene (Genbank accession number NM_153031, the representative entry for 3 underlying mRNAs also supported by EST sequences from human testis and brain libraries), only the chimp sequence had an ORF throughout the coding exon. The K_a/K_s ratio (the ratio of non-synonymous (K_a) to synonymous (K_s) nucleotide substitution rates) was 0.342, suggestive of purifying selection.

Table 1 Sequence and frequency of polymorphic mRNAs

Gene ID	Gene	BAC sequence	mRNA sequence	Numbers of individuals with genotype (out of 24)†			
				BAC/BAC	BAC/mRNA	mRNA/mRNA	n/a
<i>PMS1</i>	post-meiotic segregation increase 1	AAA*GTTACT	AAAAGTTACT	14	8	1	1
<i>PASK</i>	PAS domain-containing serine/threonine kinase	CTTTT*GCAG	CTTTT*TTTGCAG	18	5	0	1
<i>AUP1</i>	ancient ubiquitous protein 1	GTCAG*TTTT	GTCAGTTTTT	7	11	2	4
<i>TSARG2</i>	testis and spermatogenesis cell related protein 2	ATAGAGAAAT	ATA*AAAT	20	2	0	2

An asterisk in the DNA sequence indicates the position of a nucleotide difference between the corresponding BAC and mRNA sequences.

†BAC/BAC, number of individuals for whom the sequence agrees with the sequence obtained from the genomic sequencing effort; BAC/mRNA, number of heterozygotes; mRNA/mRNA, number of individuals for whom the sequence agrees with the mRNA; n/a, number of individuals for whom the sequence of the PCR product could not be obtained. See Supplementary Table 5 for additional details.

We also investigated 81 genes for which the genomic sequence differed from the corresponding mRNA, causing alterations to or truncation of the protein product predicted from the genome sequence. To determine the nature of these differences, we used polymerase chain reaction (PCR) to re-sequence each gene in a panel of 24 ethnically diverse individuals²¹ and in underlying BAC clones. Eight out of the 81 genes were found to contain errors in the genomic sequence. For 69 of the 81 genes, the sequence obtained from the 24 individuals agreed with the sequence of the original BAC clone, suggesting errors in the mRNA sequence or rare polymorphisms. Of these, 54 were single base-insertion or -deletion errors in the mRNA that shifted or truncated the reading frame in the complementary DNA relative to the genomic sequence. The other 15 mRNAs had multiple base-insertion or -deletion differences, such that the frame was eventually restored. In these cases, comparison with a related mouse gene sequence confirmed the genomic translation, typically with a more conserved match between the mouse and human genome sequences than with the corresponding human mRNA.

Four genes were determined to be polymorphic for gene disruptions in the 24 genomic samples (Table 1 and Supplementary Table 5), none of which had previously been annotated in public databases. For example, a single base-deletion event in an alternative 3' exon was observed in the *PMS1* gene, resulting in a change and extension of the ORF. Some cases of hereditary nonpolyposis colorectal cancer are associated with mutations in this gene²². In a second example, a polymorphic frameshift was observed in the final exon of the *PASK* mRNA, leading to a longer ORF in genomic translation. However, the similarity to the orthologous mouse gene does not extend past the stop codon of the mRNA. In a third case, a polymorphic single base deletion identified in the longest isoform of *AUPI* resulted in an early, in-frame stop codon, which truncated the protein at one-third of its normal length (149 out of 476 residues). Finally, an insertion of 4 bp was found in the genomic sequence of *TSARG2* (also known as *SPATA4*) relative to the RefSeq mRNA, causing a frameshift and early truncation of the protein. Similarity between the orthologous mouse protein and genomic sequence extends through the region of the frameshift.

In addition, we detected 41 potential polymorphic frameshifts by aligning genomic coding regions of chromosomes 2 and 4 (confirmed by mRNAs) to random genomic shotgun data⁹ and examining high-quality insertion and deletion differences. When comparing these regions to the orthologous chimpanzee sequence (Chimpanzee Genome Sequencing Consortium (GSC), unpublished data), the orthologous chimpanzee sequence agreed with the human BAC sequence in 37 instances. This confirms the genome sequence but does not rule out true polymorphism. For the four remaining frameshifts, confirmation with the chimpanzee sequence served to eliminate error in the random reads as a cause of the frameshift, indicating that the genome carries the derived allele. Thus, using the available chimpanzee sequence, we have identified

four probable polymorphic gene disruptions not yet annotated in the public databases.

Protein-coding genes: predicted genes

Predicted genes were identified on chromosomes 2 and 4 using GENEWISE²³ (which uses protein homologies to seed prediction), and TWINSCAN²⁴ (which uses comparative sequence analysis, in this case using the mouse genome sequence). A previous difficulty in using GENEWISE, especially in regions containing duplicated gene clusters, has been definition of the boundaries of the genomic region provided as input. We have alleviated this problem using the program BLAST2GENE²⁵ to detect (using homology) independent copies of genes, ranging from single exons to complete copies. The combined output predicted 99.5% of all known exons and at least part of 99.8% of the known genes, indicating that for known genes, the combined output is reasonably comprehensive and has high sensitivity.

Using methods similar to those for our analyses of human chromosome 7 (see ref. 8 and Supplementary Methods), we refined the initial predicted gene set. Predicted genes were required to have a highly significant match in the mouse gene set in the orthologous region of the mouse genome (if assigned), and the matching mouse gene was required to have the predicted chromosome 2 or 4 gene as its best match. Redundancy between the sets, within the sets, and with known genes was eliminated, accepting the already known genes, the GENEWISE predictions and then the TWINSCAN predictions, in that order. This yielded 314 predicted genes for chromosome 2 and 163 predicted genes on chromosome 4, bringing the total number of protein-coding genes to 1,346 and 796, respectively.

The pseudogene analysis below confirms that few functional protein-coding genes have been missed. Of the spliced ESTs mapping to chromosomes 2 and 4, 97% overlap an exon or are within 1 kb of an initial or terminal exon in the gene set. The non-overlapping ESTs might represent some coding genes, but are more likely to be other transcription products, including non-coding RNA genes or untranslated fragments of protein-coding genes. A high percentage of the predicted genes can be matched with human ESTs and have similarities with non-mammalian vertebrate protein sets (Table 2). The higher percentages for the known genes (compared with the predicted genes) probably reflect the fact that they are in general more highly expressed and, in many cases, more highly conserved across evolution. Comparisons of gene structure characteristics for known and predicted genes (Table 3) reveal similar numbers of coding bases per gene. Smaller exon counts and gene lengths for the predicted gene set might reflect fragmentation or missing terminal exons. Finally, the pseudogene analysis described below suggests that the protein-coding gene set is relatively free of pseudogenes. We conclude that the combined set of known and predicted protein coding genes is both reasonably comprehensive and free of false predictions.

Pseudogenes

With the availability of the complete genome sequence for rat²⁶ and mouse²⁷, and with the sequences for other mammalian genomes in progress, we have significantly refined our pseudogene catalogues^{8,26,28}. For chromosomes 2 and 4, we have further improved

Table 2 Coverage of predicted and known genes by various data sets

Data set	Predicted genes (%)		Known genes (%)		Ratio*	
	Chr 2	Chr 4	Chr 2	Chr 4	Chr 2	Chr 4
Human ESTs	69	58	94	93	0.73	0.62
Non-mammalian†	69	79	95	92	0.73	0.86
<i>Gallus gallus</i>	64	76	92	91	0.70	0.84
Total‡	89	92	99	99	0.90	0.93
Pfam	41	48	76	73	0.54	0.66
Interpro	48	60	82	77	0.59	0.80

Chr, chromosome.

*Ratio of predicted genes to known genes.

†Protein sets from non-mammalian vertebrate genomes: *Fugu rubripes*³⁷, *Tetraodon nigroviridis*⁴⁸, *Danio rerio* (Sanger Centre, unpublished data), and *Ciona intestinalis*⁴⁹.

‡The percentage of genes sharing similarities with human ESTs, non-mammalian vertebrate genomes and/or *Gallus gallus*.

Table 3 Characteristics of predicted versus known genes

	Predicted genes		Known genes		Ratio*	
	Chr 2	Chr 4	Chr 2	Chr 4	Chr 2	Chr 4
Exons per gene	6.6	5.3	10.5	9.5	0.63	0.56
Coding bases per gene (bp)	1,150	1,009	1,662	1,149	0.69	0.88
Genic bases per gene (kb)	33.8	34.3	66.4	75.1	0.51	0.46

*Ratio of predicted genes to known genes.

our detection and classification algorithms to define a set of 1,239 and 778 intergenic regions considered pseudogenes because they show homology to existing proteins and because nearly all ($94 \pm 3\%$ and $95 \pm 3\%$) appear to evolve neutrally according to the K_a/K_s ratio test²⁸. In contrast, only $5 \pm 3\%$ of the protein-coding set have a K_a/K_s ratio consistent with neutral evolution.

We distinguished retrotransposed (processed) from segmentally duplicated (non-processed) pseudogenes on the basis of their sequence similarity in orthologous regions of the mouse genome because processed pseudogenes tend to integrate throughout the genome (probably far from their functional paralogues). Out of 230 pseudogenes on chromosomes 2 and 4 with detectable sequence similarity to their mouse orthologous block, 94 appeared to be non-processed because they also shared similarity with neighbouring genes. Of the remaining pseudogenes (136), nearly all were shown to have lost at least one intron when compared with their parental gene, and were therefore considered processed pseudogenes. The sequence identity between these processed pseudogenes and their parental human genes is significantly higher ($>10\%$ difference in nearly all cases) than the sequence identity between the pseudogenes and the orthologous mouse regions. This strongly suggests that the pseudogenes arose independently in the human and mouse lineages, rather than from a common ancestor as recently proposed²⁹. Altogether, we defined a set of 1,856 processed and 161 non-processed pseudogenes, consistent with estimates of $\sim 20,000$ pseudogenes across the human genome²⁸.

Non-coding RNAs

We identified non-coding RNA (ncRNA) genes in the chromosome

sequences as described for the human genome sequence¹. With the exception of close sequence homologues and transfer RNAs, ncRNA gene annotation is still largely limited to annotating known gene sequences³⁰. Chromosomes 2 and 4 contain 15% of the unambiguous bases in the human genome and 14% of the annotated ncRNA pseudogenes (863 out of 6,124), but only 2% of the tRNAs and only 5% of all annotated ncRNA genes (50 out of 1,096). The ncRNA pseudogene annotation includes 163 tRNA pseudogenes, 150 of which were derived from ancient mtDNA integrations into the nuclear genome (nuclear mitochondrial DNAs or NUMTs³¹). Previous annotation missed most of these mtDNA-derived pseudogenes, because their sequence divergence is usually too great to recognize them individually. However, here we have identified these pseudogenes by searching for clustering of weak hits in a larger expanse of decayed mtDNA synteny.

Protein index

We derived an index of predicted protein sequences for human chromosomes 2 and 4, and compared them to the Interpro database³² using Interproscan³³ to predict protein families, domain and repeat families, and sequence motifs. Of the 74% of proteins that had an Interpro classification, 67% were multi-domain. Protein kinases are the most highly represented families on the two chromosomes. Gene clusters included the most prevalent protein families in the human genome, the immunoglobulins and zinc-finger domain-containing proteins. On chromosome 2, a cluster of 13 genes containing immunoglobulin-like domains (Interpro identifier IPR007110) is found on the p-arm near the centromere, and a cluster of 11 genes containing zinc-finger domains (C_2H_2

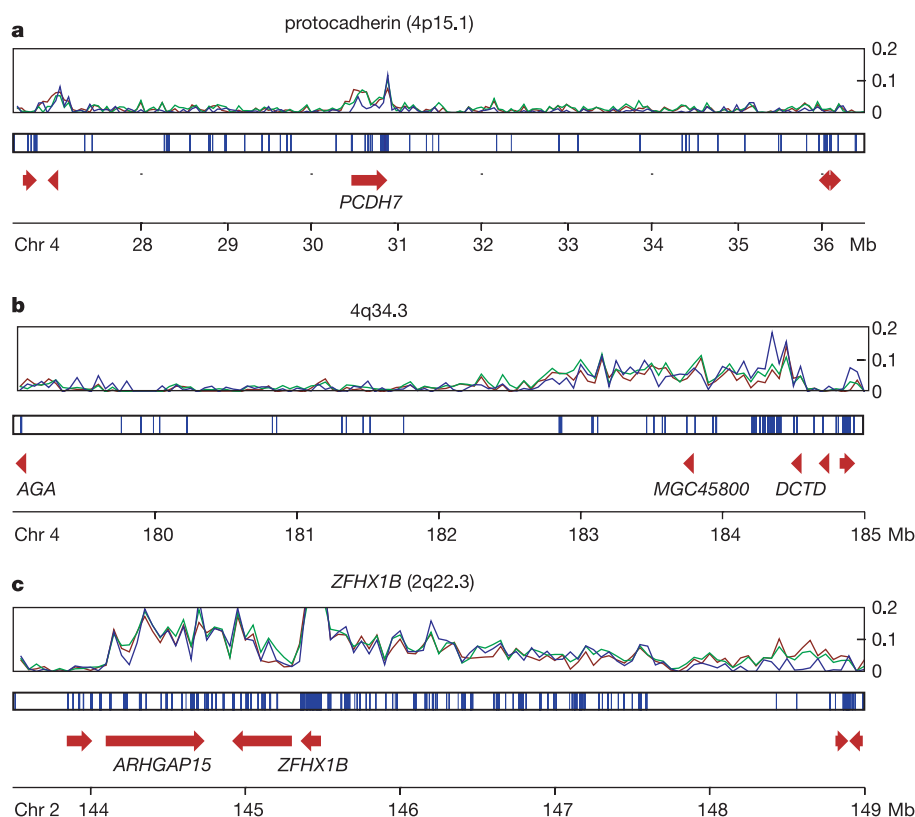


Figure 1 Three gene deserts. Levels of non-coding conservation are plotted in non-overlapping 50-kb windows between human and dog (green), human and mouse (red), and human and chicken (blue). The second panel in each of **a–c** shows positions of matches between human and *Fugu rubripes* (including coding regions) computed by the BLAT⁵⁰ program. Red symbols show the positions and orientations of genes. **a**, A 10-Mb

interval of chromosome 4 centred on the protocadherin gene *PCDH7*. **b**, A 6-Mb interval of chromosome 4 containing one of the longest human gene deserts, showing significant enrichment for matches to chicken, mouse and dog. **c**, A 5.5-Mb interval of chromosome 2 that contains the longest human segment showing high enrichment for non-coding matches with chicken.

type; IPR007087) is found on the q-arm. On chromosome 4, the largest clusters are grouped near each other on the q-arm: one cluster contains nine genes with UDP-glucosyltransferase domains (IPR002213), and another cluster contains eight small chemokines (of the C-X-C/interleukin-8 subfamily; IPR002473). The Interpro results were used to assign Gene Ontology (GO) codes³⁴: for chromosomes 2 and 4, 63% and 68% of proteins were assigned to the global category of molecular function, 51% and 57% to biological process, and 34% and 39% to cellular components, respectively. The two most frequent specific GO categories for both chromosomes were the cellular components of nucleus (GO:0005634) and membrane (GO:0016020), representing 15% (chromosome 2) and 19% (chromosome 4) of the genes with assigned function.

Gene deserts and conserved non-coding sequence

Gene deserts are a curious feature of vertebrate genomes; they are megabase-size genomic segments devoid of protein-coding genes. The overall architectures of these regions are maintained over long evolutionary distances, sometimes with only very limited sequence conservation³⁵. In other cases, the deserts have been found to contain small blocks of highly conserved sequences that regulate flanking genes. Roughly 80% of human gene deserts occur in (G+C)-poor, Giemsa-dark chromosome bands. Chromosomes 2 and 4, which contain some of the largest deserts in the genome, provide opportunities to examine possible roles of such regions.

Strikingly, *PCDH7*, a protocadherin gene expressed predominantly in the brain and heart (Fig. 1a), and its paralogue *PCDH10* (at 4q28.3), are each flanked on both sides by unusually large deserts (5.2 Mb and 3.5 Mb for *PCDH7*, and 5.1 Mb and 4.0 Mb for *PCDH10*). The overall genome architecture of these gene deserts, including the flanking genes, is conserved in mammals and birds (data not shown), but the deserts themselves show only average levels of nucleotide conservation with dog (data provided by the Broad Institute), mouse²⁷ and chicken³⁶. The duplication event that separated the two protocadherins occurred before mammals diverged from fish, but after they diverged from *Ciona intestinalis*, suggesting that this arrangement has persisted for hundreds of millions of years.

Two other large deserts (Fig. 1b, c) have portions with a higher-than-average level of nucleotide conservation; these are a 4.7 Mb desert at 4q34.3 and a 3.5 Mb desert upstream of *ZFH1B* at 2q22.3. When compared with chicken, we found 168 segments of conserved sequence for the desert at 4q34.3 (36.5 per Mb), 246 segments for the desert at 2q22.3 (74.5 per Mb), and 64 segments in the desert

downstream of *PCDH7* (13.1 per Mb). Within these intervals, the conserved segments are not uniformly distributed. For the 4q34.3 desert, the pattern of conservation is similar to that seen in mouse and dog, but there is relatively little conservation with the pufferfish *Takifugu (Fugu) rubripes*³⁷. In contrast, the region upstream of *ZFH1B* is enriched in conserved segments in *T. rubripes* as well as in dog and mouse. The *ZFH1B* gene and its neighbour *ARHGAP15* appear to be products of a segmental duplication. Their paralogous copies (*TCF8* and *ARHGAP12*) are found on 10p11.22. The intergenic interval upstream of *TCF8* is 335 kb and shows only weak interspecies conservation. There is also little evidence for conservation between the paralogues outside of the coding regions, other than a weak alignment in the 3'-UTRs of *ZFH1B* and *TCF8*. This UTR region is almost 100% conserved between *ZFH1B* and the orthologous intervals in chicken and mouse. Thus, the presence of these gene deserts is maintained over longer evolutionary periods than are intervals of high nucleotide-level conservation within them.

We looked for overrepresented motifs in the conserved non-coding segments in the 2q22.3 region. We evaluated the frequencies of short (4–9 bp) patterns in the conserved and diverged intervals (see Supplementary Methods for details). Using these patterns, we could distinguish between the conserved and non-conserved intervals with 75% accuracy, suggesting that the conserved regions share short, specific, non-coding (but presumably functional) elements, perhaps transcription factor binding sites, that have been conserved throughout the evolution of mammals and birds.

Recent local duplications—complexity of gene prediction

The gene for a well-characterized human nuclear export protein, the RAN binding protein 2 (*RANBP2*), is contained in a region of recent duplication in the human genome and is in a recombination hot spot on chromosome 2q (ref. 38). Through a series of rearrangements, this region underwent exon shuffling, domain accretion and deletion, suggesting that this entire region has been extremely dynamic over the last several million years. In a detailed examination of this region, we discovered eight new genes that arose by duplication of *RANBP2* (Fig. 2). We called the new gene family *RGP*, for *RANBP2*-like, GRIP domain-containing proteins. The *RGP* copies are interspersed in a 9-Mb region on both sides of the centromere, and have significantly modified their gene structure compared with *RANBP2*. There is experimental evidence for expression of almost all of the copies³⁹, including specific ESTs and cDNAs. Similar duplications are also found in chimpanzee (but other sequenced metazoan genomes, including mouse, have only

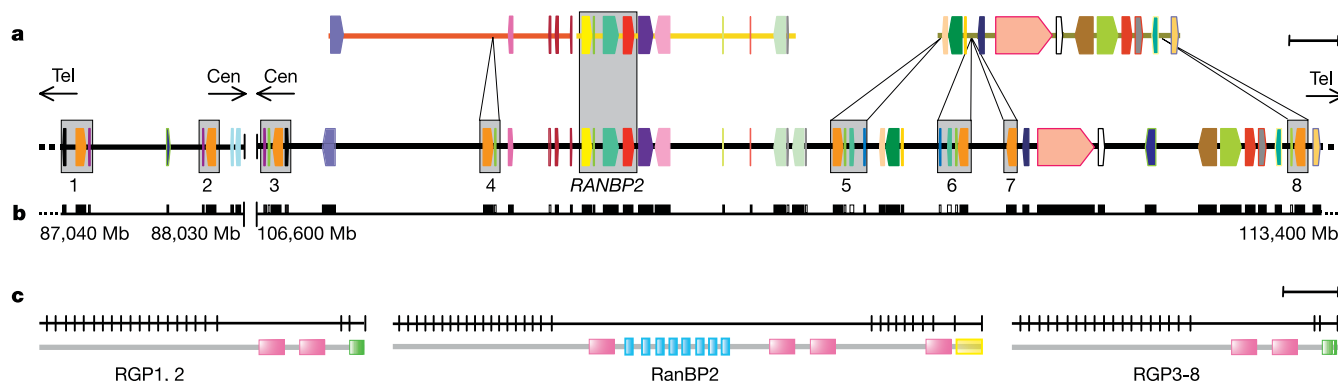


Figure 2 *RANBP2* gene duplication in the pericentromeric region of chromosome 2. **a**, Gene composition. Each coloured bar/arrow represents a different gene; segmentally duplicated regions are highlighted with grey boxes. *RANBP2* and its eight paralogues are shown as red and orange arrows, respectively. Regions of orthology in mouse chromosomes 17 (orange), 10 (yellow) and 2 (green) are reported above the human gene

bar. Scale bar, 310 kb. **b**, Filled boxes indicate evidence for expression of genes in **a**. **c**, Gene and protein domain architecture of *RANBP2* and the *RGP* family (*RGP1*–*8*). Black lines show exon/intron boundaries; pink boxes show Ran binding domains; blue boxes show zinc-finger domains; yellow box shows a cyclophilin A-homologous domain; green boxes show GRIP domains. Scale bar, 300 amino acids.

one copy). This region provides an excellent example of difficulties encountered during annotation using automated pipelines, and shows how the application of new tools for hunting for gene duplications can reveal interesting evolutionary scenarios.

Segmental duplications

Segmental duplications of genomic DNA—large, low-copy repeats often spanning hundreds of kilobases—are a prominent feature of the human genome. We performed a detailed analysis of duplicated sequence ($\geq 90\%$ sequence identity and ≥ 1 kb in length), comparing the finished chromosome assemblies to the human genome sequence. Both chromosomes 2 and 4 show less segmental duplication (4.2% and 2.3%, respectively) than the genome average (5.2%) (Supplementary Table 6 and Supplementary Fig. 2). The reduction in segmental duplications on chromosome 4 is especially noticeable within the pericentromeric regions of 4q11–q12, where not a single duplication could be detected within 2 Mb of the centromere. Although it remains possible that additional pericentromeric sequence will be recovered for 4q11–q12, the most proximal segment of sequence does contain ~ 26 kb of alpha monomeric satellite DNA, suggesting proximity to the centromere. Given that we found no evidence for duplications using a second detection strategy⁴⁰, and given that fluorescence *in situ* hybridization (FISH) data⁴¹ provide little evidence for 4q11–q12 pericentromeric duplications, we suggest that 4q11–q12 has been relatively quiescent in terms of pericentromeric duplications. This is in sharp contrast with 2p11 and 2q11, where nearly half of the most proximal 2-Mb pericentromeric region (899 kb and 910 kb respectively) shows extensive recent duplication (Supplementary Fig. 3).

Chromosome 2 is unique to the human lineage of evolution, having emerged as a result of head-to-head fusion of two acrocentric chromosomes that remained separate in other primates. The precise fusion site has been located in 2q13–2q14.1 (ref. 2;

hg16:114455823–114455838), where our analysis confirmed the presence of multiple subtelomeric duplications to chromosomes 1, 5, 8, 9, 10, 12, 19, 21 and 22 (Fig. 3; Supplementary Fig. 3a, region A). During the formation of human chromosome 2, one of the two centromeres became inactivated (2q21, which corresponds to the centromere from chimp chromosome 13) and the centromeric structure quickly deteriorated⁴². A search of genome sequence for the presence of vestigial centromere and pericentromeric sequences identified a 2.6-Mb region in 2q21.1–2q21.2 that is enriched for pericentromeric duplications to chromosomes 1, 7, 9, 10, 13, 14, 15, 18, 21 and 22 as well as a variety of centromeric satellite repeat sequence motifs (HSAT5, GSATII, ACRO1). The degree of sequence identity of the interchromosomal duplications ($< 98\%$) suggests that these pericentromeric segmental duplications existed before the formation of this chromosome. Within this 2.6-Mb interval, we identified a relatively large tract of satellite sequence (three tracts totalling 31,198 bp of alpha-satellite sequence over 36,696 bp), which likely demarcates the position of the ancestral centromere (Supplementary Fig. 3a, region B). These data raise the possibility that ancestral telomeres and ancestral centromeres that have disappeared over the course of mammalian chromosomal evolution might be marked by the presence of an abundance of residual pericentromeric and subtelomeric duplications.

By analogy, an interstitial 1.1-Mb region of subtelomeric duplications was identified within 4q26 (Supplementary Fig. 3b, region D). This region probably represents a genomic segment that became duplicately transposed to a subtelomeric region and was subsequently dispersed throughout the human genome by secondary subtelomeric–subtelomeric duplications. The average percentage identity between this region and the subtelomeric regions is 97.4%, and among these subtelomeric regions it is 99.0%, indicating more recent duplications or gene conversion events. Although there is not conclusive evidence that this

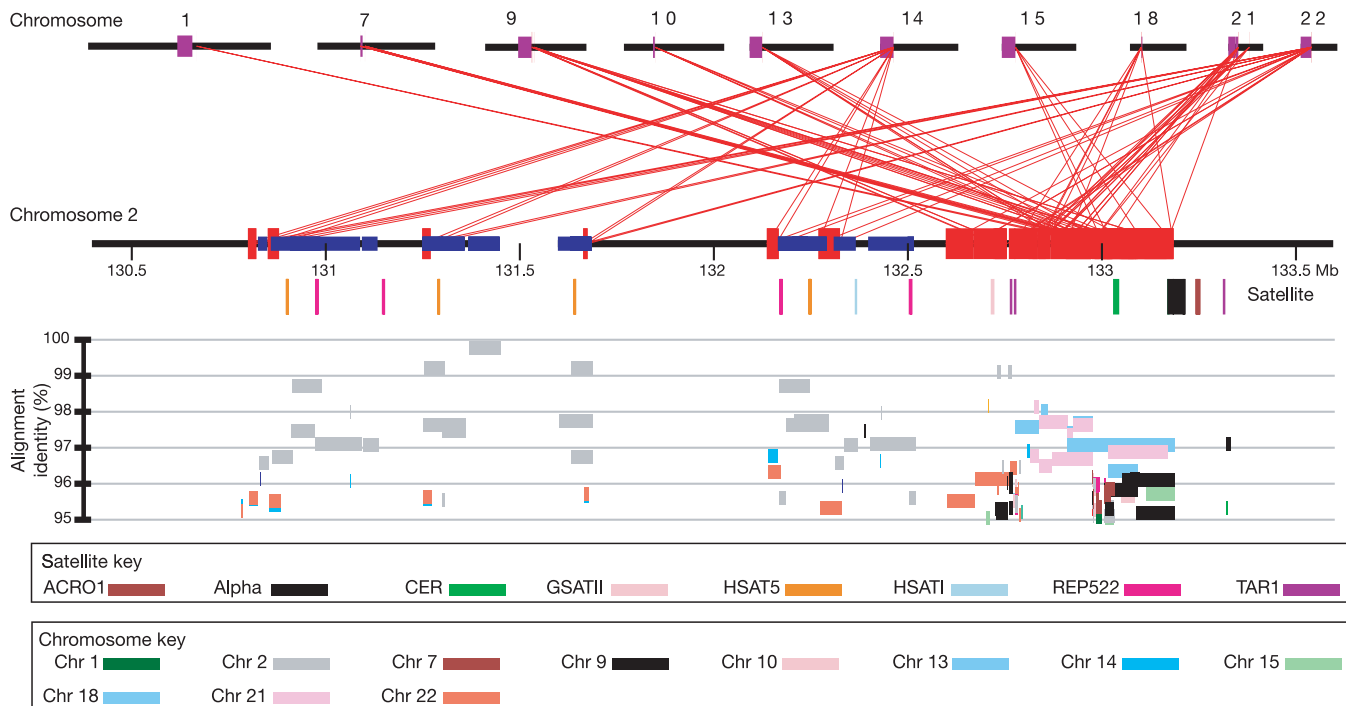


Figure 3 Pattern of recent segmental duplications in the vestigial centromere region on chromosome 2. Large (> 10 kb), highly similar ($> 95\%$) intrachromosomal (blue) and interchromosomal (red) segmental duplications are shown for a 3-Mb region along the horizontal line of chromosome 2, in increments of 0.5 Mb. The upper panel shows

extensive duplication with pericentromeric regions on other chromosomes. Centromeres are shown in purple. The coloured vertical bars underneath chromosome 2 are satellite sequences. The lower panel shows the percentage identity of each pairwise alignment. Coloured bars represent alignments from different chromosomes.

represents the site of an ancestral telomere, it has the characteristics of such a region, showing an abundance of subtelomeric duplications to human chromosomes 1–9, 11, 16, 19 and 20, an enrichment of short blocks of orthology between human and mouse, and a breakpoint of conserved orthology with the flanking regions between human and mouse. A second short (30-kb) region in 4q32.3 also shows a breakpoint in conservation with mouse and a similar pattern of subtelomeric duplications. Sixteen duplications ranging in size from 10–20 kb and showing remarkably high sequence identity (96–97%) are distributed among ten subtelomeric regions (chromosomes 1, 2, 5–8, 10, 16, 19 and 20, Supplementary Fig. 3b, region E). In contrast with the region of 4q26, the organization of this region appears to be conserved in mouse, as revealed by orthologous mouse–human anchor sequences that extend beyond the duplication. Finally, although intrachromosomal duplications are relatively rare on chromosome 4 (Supplementary Figs 4 and 5), one large (750-kb) cluster of tandem intrachromosomal duplication was noted in 4q13.2. This region contains at least five members of a family of tandemly duplicated microsomal UDP-glycosyltransferase genes that are thought to be important in drug detoxification.

Sequence variation

In the process of constructing contiguous chromosomal sequences from the underlying BAC clone sequences, we encountered several candidate overlaps that showed unusually high levels of variation. For 53 overlaps, totalling 3.1 Mb (with an average difference of ~4 differences per kb), we showed by segregation analysis⁸ that these represented true overlaps, with the differences arising from different alleles of the same locus and not from distinct regions produced by segmental duplication²¹. Extending this analysis to 2,718 overlap regions on chromosomes 2, 4 and 7 with at least 5 kb of overlap, we identified 678 regions with at least three ‘polymorphic events’, suggesting that the underlying clones represented different haplotypes⁴³. Of greatest interest were 24 overlaps (Supplementary Table 7), where multiple (≥ 3) consecutive windows varied by at least two standard deviations from the average, possibly indicating segments showing balancing selection⁴⁴. In one example, a 5-kb segment had 75 polymorphic events, and the neighbouring five windows also had more than 18 (2 s.d.) polymorphic events. In 15 cases, the highly variant region was found near a gene, including the known genes *CRYPTIC* (also known as *CFC1*), *TSSC1*, *LRP1B* and *GYPE*. We sampled across eight of these highly variant regions from the panel of 24 individuals²¹, and confirmed both that these segments were polymorphic in the population and that these regions of extreme variation appear to arise from two distinct haplotype blocks. Further, chimp single nucleotide polymorphisms (SNPs) mapped to the human genome (Chimpanzee GSC, unpublished data) show increased frequencies of polymorphisms (at least four consecutive 5-kb windows containing more than mean + 2 s.d.) in three of the five regions where orthologous chimpanzee sequence was available. Based on these data, it is unlikely that this represents a locally high mutation rate. Instead, these regions might represent unusually deep coalescents, perhaps as a result of diversifying selection.

Summary and conclusions

The achievement of highly accurate comprehensive sequence for chromosomes 2 and 4 (and other published human chromosomes) represents a critical step for the Human Genome Project in that it makes possible more detailed and conclusive analyses. On chromosome 2, the local region surrounding the ancestral chromosomal fusion site on 2q13–2q14.1 had previously been described². Here we identified a 2.6-Mb region within 2q21.1–2q21.2 that is enriched for pericentromeric duplications and centromeric satellite repeat sequence motifs, including a stretch of alpha-satellite sequence that probably identifies the location of the ancestral centromere.

Our results suggest that the abundance of residual pericentromeric and subtelomeric duplications in the human genome, coupled with breakpoints in orthology with mouse, might be markers of ancestral telomeres and centromeres, perhaps representative of recurrent chromosome rearrangements in the human lineage. Also, using orthologous relationships with other genome sequences such as mouse and rat, we have more confidently identified and discriminated between genes and pseudogenes, with accurate classification regarding processed and non-processed pseudogenes. Although the expanding mRNA sets still require cautious evaluation, we have used them to identify a possible human/chimpanzee-specific protein and to characterize a small subset of proteins that contain polymorphisms leading to altered proteins in the human lineage. The sequence data for chromosomes 2 and 4, together with the high-quality annotation of genes and an analysis of homologous regions of vertebrate genomes, reveal a high-level structure consisting of two gene deserts separated by a protocadherin gene. This high-level structure appears to be more stable over evolutionary time than does the low-level nucleotide conservation in other deserts. A genome-wide search revealed a general property of the cadherin gene family: it includes 12% of all human genes that have a desert immediately on each side, but contains only 0.4% of all human genes. We have also described genomic segments (on the order of tens of kilobases) for which there are sharp rises in the density of human variation that seem to correlate with haplotype blocks (for example, ref. 43). Although these regions might still represent chance preservation of two alternate haplotype blocks over evolutionary time, they might well also be the result of balancing selection in these regions. It will be important to continue evaluating additional regions, their possible roles in human variation and their correlations with human phenotypes. Critically, immediate access to this sequence will allow researchers in all disciplines to contribute to the understanding of the genome and its relevance in human health. □

Methods

Sequencing in other primates

We attempted to amplify across each exon in each gene of interest from the following primates: *Homo sapiens*, Celebes crested macaca (*Macaca nigra*), Sumatran orangutan (*Pongo pygmaeus*), *Gorilla gorilla*, black-handed spider monkey (New World monkey *Ateles geoffroyi*) and chimpanzee (*Pan troglodytes*). Primers were chosen in highly conserved human/chimp intronic regions directly flanking the exons. Multiple primers were chosen to increase the possibility of getting a successful product. When we were not able to amplify a given exon in a primate, new primers were chosen on the basis of sequence conservation data from other monkeys where amplification had been successful.

Assessing human deletions/polymorphisms

A list of possible deletions or polymorphisms in human chromosomes 2 and 4 was created by placement of the fosmid end sequences against the human genome⁴. All fosmid end placements separated by less than 3.5 standard deviations from the average were flagged as possible deletions. Using those criteria, 27 possible deletion regions were flagged (Supplementary Tables 8 and 9). Almost 70% of these regions across the human genome were found to be polymorphic when tested using PCR analysis⁴. We aligned both of the publicly available *Pan troglodytes* assemblies (Chimpanzee GSC, unpublished data) against the human genome and manually reviewed the alignments to see whether the possible deletions in human were confirmed by chimpanzee coverage. In 11 cases (41%), there was good coverage of the entire region in the chimpanzee assembly and there was no suggestion of missing data in the human genome. In 12 cases (44%), there was additional chimpanzee data throughout the region, suggesting a possible deletion (polymorphic or real) in the human sequence. In one case, there was no chimp coverage of the human region, and in three cases the chimpanzee assembly was too fragmented to determine whether a deletion was present. It remains to be determined whether the deletion in the human sequence was an error or polymorphic.

Gene deserts

For each panel of Fig. 1, we identified orthologous intervals in the current (as of July 2004) dog, mouse and chicken genome assemblies. After masking known coding regions in the human sequence, we performed alignments to the other three species using the BlastZ program⁴⁵. For the human–chicken alignment, default alignment-scoring parameters were used. For human–dog alignment, we used the following scores: match = 100; transition = –500; transversion = –1,200; gap-open = –2,000; and gap-extension = –500. For human–mouse the respective parameters were 100, –200, –600, –1,000 and –200. Those parameter values were chosen so that the genome-wide

fraction of aligned bases was about 0.02 for each species; segments with a higher fraction of aligned bases were considered to be enriched for interspecies conservation. Figure 1 shows the fraction of the segment contained within a local alignment, for non-overlapping 50-kb windows.

Classification of conserved versus nonconserved segments of the human gene desert in Fig. 1c was performed as follows. The gene desert was aligned to the orthologous chicken region, and human intervals showing strong conservation (using an arbitrary threshold of at least 70% identity with chicken over 100 bp of sequence) or no conservation with chicken were selected. To ensure equal training-set size and length distribution, a subset of the unaligned intervals was prepared by repeatedly randomly selecting an element from the strongly aligned set, and then randomly selecting a piece of an interval from the unaligned set of the same size.

Classification was first performed using the Markov model method described in ref. 46, using only the human sequence (rather than an alignment) as input and the four possible nucleotides as an alphabet. Leave-one-out cross-validation was performed for each pair of training sets at orders 1 through 8 (that is, nucleotide words of 2–9 bp) to determine the best classifying model.

The alternative classification approach constructed a vector of the frequency of occurrence of each possible word for each interval. A support vector machine (SVM) as implemented in LIBSVM⁴⁷ was used to classify each pair of training sets. A gaussian kernel was used and parameters were selected using a grid search and fivefold cross-validation. Because the grid search is more expensive, only word sizes 4–6 were tested.

For comparison, we applied the same methods to sets of RefSeq-annotated 3' - and 5' -UTRs with similar (G+C) content and length distributions. The 3' - and 5' -UTR sets were chosen by selecting the subset of all RefSeq-annotated UTRs with (G+C) content between 0.36 and 0.37 (similar to the gene desert) and length between 20 and 400 bp (resulting in a similar length distribution). We were able to distinguish them with 67% accuracy. This suggests that the human sequence of the strongly conserved and unconserved intervals in this gene desert show at least as much difference in characteristic short patterns used by the classification schemes as do 3' - and 5' -UTRs.

Received 25 October 2004; accepted 11 February 2005; doi:10.1038/nature03466.

1. Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
2. Fan, Y., Linardopoulou, E., Friedman, C., Williams, E. & Trask, B. J. Genomic structure and evolution of the ancestral chromosome fusion site in 2q13–2q14.1 and paralogous regions on other human chromosomes. *Genome Res.* **12**, 1651–1662 (2002).
3. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
4. The International Human Genome Mapping Consortium. A physical map of the human genome. *Nature* **409**, 934–941 (2001).
5. Felsenfeld, A., Peterson, J., Schloss, J. & Guyer, M. Assessing the quality of the DNA sequence from the Human Genome Project. *Genome Res.* **9**, 1–4 (1999).
6. Schmutz, J., Wheeler, J., Grimwood, J., Dickson, M. & Myers, R. M. Assessing the quality of finished genomic sequence. *Cold Spring Harb. Symp. Quant. Biol.* **68**, 31–37 (2003).
7. Riethman, H. C. *et al.* Integration of telomere sequences with the draft human genome sequence. *Nature* **409**, 948–951 (2001).
8. Hillier, L. W. *et al.* The DNA sequence of human chromosome 7. *Nature* **424**, 157–164 (2003).
9. The International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
10. Dib, C. *et al.* A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**, 152–154 (1996).
11. Broman, K. W., Murray, J. C., Sheffield, V. C., White, R. L. & Weber, J. L. Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am. J. Hum. Genet.* **63**, 861–869 (1998).
12. Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nature Genet.* **31**, 241–247 (2002).
13. Schuler, G. D. *et al.* A gene map of the human genome. *Science* **274**, 540–546 (1996).
14. Olivier, M. *et al.* A high-resolution radiation hybrid map of the human genome draft sequence. *Science* **291**, 1298–1302 (2001).
15. Antequera, F. & Bird, A. Number of CpG islands and genes in human and mouse. *Proc. Natl Acad. Sci. USA* **90**, 11995–11999 (1993).
16. Wang, Y. & Leung, F. C. An evaluation of a new criteria for CpG islands in the human genome. *Bioinformatics* **20**, 1170–1177 (2004).
17. Liu, C. X., Musco, S., Lisitsina, N. M., Yaklichkin, S. Y. & Lisitsyn, N. A. Genomic organization of a new candidate tumor suppressor gene, LRP1B. *Genomics* **69**, 271–274 (2000).
18. Chan, A. C. *et al.* ZAP-70 deficiency in an autosomal recessive form of severe combined immunodeficiency. *Science* **264**, 1559–1601 (1994).
19. Pruitt, K. D. & Maglott, D. R. RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Res.* **29**, 137–140 (2001).
20. Mammalian Gene Collection Program Team. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc. Natl Acad. Sci. USA* **99**, 16899–16903 (2002).
21. Sachidanandam, R. *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933 (2001).
22. Nicolaides, N. C. *et al.* Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. *Nature* **371**, 75–80 (1994).
23. Birney, E. & Durbin, R. Using GeneWise in the *Drosophila* annotation experiment. *Genome Res.* **10**, 547–548 (2000).
24. Korf, I., Flicek, P., Duan, D. & Brent, M. R. Integrating genomic homology into gene structure prediction. *Bioinformatics* **17**(S1), 140–148 (2001).

25. Suyama, M., Torrents, D. & Bork, P. BLAST2GENE: a comprehensive conversion of BLAST output into independent genes and gene fragments. *Bioinformatics* **20**, 1968–1970 (2004).
26. Rat Genome Sequencing Project Consortium. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**, 493–521 (2004).
27. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
28. Torrents, D., Suyama, M., Zdobnov, E. & Bork, P. A genome-wide survey of human pseudogenes. *Genome Res.* **13**, 2559–2567 (2003).
29. Zhang, Z., Carriero, N. & Gerstein, M. Comparative analysis of processed pseudogenes in the mouse and human genomes. *Trends Genet.* **20**, 62–67 (2004).
30. Eddy, S. R. Computational genomics of noncoding RNA genes. *Cell* **109**, 137–140 (2002).
31. Mishmar, D., Ruiz-Pesini, E., Brandon, M. & Wallace, D. C. Mitochondrial DNA-like sequences in the nucleus (NUMTs): Insights into our African origins and the mechanism of foreign DNA integration. *Hum. Mutat.* **23**, 125–133 (2004).
32. Mulder, N. J. *et al.* InterPro: an integrated documentation resource for protein families, domains and functional sites. *Brief. Bioinform.* **3**, 225–235 (2002).
33. Zdobnov, E. M. & Apweiler, R. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848 (2001).
34. The Gene Ontology Consortium. Creating the Gene Ontology resource: design and implementation. *Genome Res.* **11**, 1425–1433 (2001).
35. Ovcharenko, I. *et al.* Evolution and functional classification of vertebrate gene deserts. *Genome Res.* **15**, 137–145 (2005).
36. International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–715 (2004).
37. Aparicio, S. *et al.* Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* **297**, 1301–1310 (2002).
38. Bailey, J. A. *et al.* Recent segmental duplications in the human genome. *Science* **297**, 1003–1007 (2002).
39. Ciccarelli, F. D. *et al.* Complex genomic rearrangements lead to novel primate gene function. *Genome Res.* **15**, 343–351 (2005).
40. Bailey, J. A. *et al.* Human-specific duplication and mosaic transcripts: the recent paralogous structure of chromosome 22. *Am. J. Hum. Genet.* **70**, 83–100 (2002).
41. The BAC Resource Consortium. Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature* **409**, 953–958 (2001).
42. Yunis, J. J. & Prakash, O. The origin of man: a chromosomal pictorial legacy. *Science* **215**, 1525–1530 (1982).
43. Gabriel, S. B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).
44. Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J. & Lander, E. S. High-resolution haplotype structure in the human genome. *Nature Genet.* **29**, 229–232 (2001).
45. Schwartz, S. *et al.* Human-mouse alignments with BLASTZ. *Genome Res.* **13**, 103–107 (2003).
46. Kolbe, D. *et al.* Regulatory potential scores from genome-wide three-way alignments of human, mouse, and rat. *Genome Res.* **14**, 700–707 (2004).
47. Chang, C. C. & Chih-Jen, L. LIBSVM: a library for support vector machines. Software at (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>) (2001).
48. Roest-Crolius, H. *et al.* Estimate of human gene number provided by genome-wide analysis using *Tetradon nigroviridis* DNA sequence. *Nature Genet.* **25**, 235–238 (2000).
49. Dehal, P. *et al.* The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* **298**, 2157–2167 (2002).
50. Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements The authors would like to thank the staff, past and present, of the Washington University Genome Sequencing Center as well as the following laboratories that contributed sequence segments to the final chromosome 2 and 4 sequence: the Broad Institute of MIT; Stanford DNA Sequencing and Technology Development Center; the Sanger Center; National Yang-Ming University in Taipei; Genoscope; Baylor College of Medicine; the Joint Genome Institute; University of Washington Multimegabase Sequencing Center; DOE Joint Genome Institute; and Roswell Park Cancer Institute. We thank T. Hubbard, J. Ashurst, J. Gilbert and S. Keenan for Ensembl pipeline assistance. We thank the Broad Institute of MIT, Harvard and Agencourt Bioscience for making a preliminary assembly of dog genome sequence available to us. We thank the Zebrafish Sequencing Group at the Sanger Centre for the *Danio rerio* preliminary protein set. We thank the International HapMap Consortium and their sponsors for use of their data and J. Mullikin for help with alignment of their data to the human genome. We thank the Chimpanzee Genome Sequencing Consortium for use of the chimpanzee sequence data. The authors would also like to acknowledge the efforts of the HUGO Gene Nomenclature Committee. Finally, the authors acknowledge the National Human Genome Research Institute for funding this work.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.K.W. (rwilson@watson.wustl.edu). All reported DNA sequences have been deposited in GenBank or EMBL. Accession numbers for the chromosome sequence analysed for this paper can be found in Supplementary Table 1. The updated chromosome 2 and 4 sequences can be accessed through GenBank accession numbers NC_000002 (chromosome 2) and NC_000004 (chromosome 4). Primate resequencing data can be accessed using GenBank accession numbers CZ179368–CZ179565. The mRNA resequencing data can be accessed via GenBank/dbSNP identifiers ss35032449–ss35032461, ss35033273–ss35033317.

Recruitment of entomopathogenic nematodes by insect-damaged maize roots

Sergio Rasmann¹, Tobias G. Köllner², Jörg Degenhardt², Ivan Hiltbold¹, Stefan Toepfer³, Ulrich Kuhlmann³, Jonathan Gershenzon² & Ted C. J. Turlings¹

¹University of Neuchâtel, Institute of Zoology, Laboratory of Animal Ecology and Entomology, CP 2, CH-2007 Neuchâtel, Switzerland

²Max Planck Institute for Chemical Ecology, Hans-Knöll-Strasse 8, D-07745 Jena, Germany

³CABI Bioscience Switzerland Centre, Rue des Grillons 1, 2800 Delémont, Switzerland

Plants under attack by arthropod herbivores often emit volatile compounds from their leaves that attract natural enemies of the herbivores. Here we report the first identification of an insect-induced belowground plant signal, (*E*)- β -caryophyllene, which strongly attracts an entomopathogenic nematode. Maize roots release this sesquiterpene in response to feeding by larvae of the beetle *Diabrotica virgifera virgifera*, a maize pest that is currently invading Europe. Most North American maize lines do not release (*E*)- β -caryophyllene, whereas European lines and the wild maize ancestor, teosinte, readily do so in response to *D. v. virgifera* attack. This difference was consistent with striking differences in the attractiveness of representative lines in the laboratory. Field experiments showed a fivefold higher nematode infection rate of *D. v. virgifera* larvae on a maize variety that produces the signal than on a variety that does not, whereas spiking the soil near the latter variety with authentic (*E*)- β -caryophyllene decreased the emergence of adult *D. v. virgifera* to less than half. North American maize lines must have lost the signal during the breeding process. Development of new varieties that release the attractant in adequate amounts should help enhance the efficacy of nematodes as biological control agents against root pests like *D. v. virgifera*.

Plants are not simply passive victims of attacking herbivores; they have evolved an arsenal of physical and chemical defences to protect themselves. Often these defences are mobilized only in response to herbivory^{1,2}. Among the proposed inducible defences is the production and release of volatile chemicals that could serve as signals to attract natural enemies of the herbivores^{3–5}. Manipulating these signals can help increase the effectiveness of these natural enemies as control agents^{6–8}. The induced emission of chemical signals is not limited solely to aboveground plant parts. The entomopathogenic nematode *Heterorhabditis megidis* was found to be attracted to exudates emitted by plant roots after damage by weevil larvae^{9,10}, but the nature of the attractants involved is unknown. Here we show that maize roots damaged by larvae of the economically important coleopteran pest *Diabrotica virgifera virgifera* LeConte are attractive to entomopathogenic nematodes, and we identify the chemical compound responsible for the attraction. *D. v. virgifera* or Western corn rootworm (WCR) is a voracious pest of maize that is responsible for the use of the bulk of pesticides applied in the cultivation of this crop in the USA¹¹. The recent introduction and rapid spread of WCR into Europe has caused major concern for maize production on this continent and has stimulated the search for new methods of maize protection^{12,13}. The use of nematodes to control WCR is an ecologically sound option^{14,15}, especially if researchers can optimize their efficacy at finding and killing WCR.

Attraction of nematodes by WCR-damaged roots

To determine whether or not WCR-infested maize plants would attract nematodes, three glass pots each containing one 10-day-old maize plant (var. Delprim) were attached to the arms of a custom-made six-arm olfactometer filled with moist (10% water) sand (Fig. 1a). The plants had been grown on clean sand in the pots, starting 5 days after seed germination. Three additional pots, containing only sand, were attached to the remaining three arms

of the olfactometer. Four such olfactometers, each containing three plants plus three sand controls, were prepared on a given day. One plant of each set of three received four second-instar or third-instar WCR larvae, the roots of a second plant were damaged daily by stabbing them five times with a metal corkborer 7 mm in diameter, and the third plant was left unharmed. On day 3 after initial damage, about 2,000 *Heterorhabditis megidis* nematodes were released in the centre of each olfactometer, where they were free to enter the arms until their passage was blocked by an ultra-fine metal screen (see Fig. 1a and Methods). One day after release, the number of nematodes in each arm was recorded. Significantly more nematodes were recovered from arms connected to the pots with the WCR-damaged plants than from the arms connected to the other treatments or controls (Fig. 1b), indicating that damage by WCR induces maize roots to release a nematode attractant.

Identification of the attractant

Maize leaves had previously been shown to emit a mixture of volatile compounds in response to damage by caterpillars⁴. To determine whether WCR damage induces similar changes in plant volatiles, the leaves and roots from WCR-damaged (3 days) and healthy maize plants were ground and volatiles collected by solid-phase micro-extraction (SPME) were analysed by gas chromatography–mass spectrometry (GC–MS). A marked difference between the treatments was that the sesquiterpene (*E*)- β -caryophyllene was present in roots damaged by WCR but was completely absent from undamaged roots (Fig. 1c). The damaged roots contained small amounts of α -humulene and caryophyllene oxide as well. To a smaller extent, the WCR-induced increase in (*E*)- β -caryophyllene content was also apparent in the leaves (Fig. 1d). To test whether (*E*)- β -caryophyllene was indeed attractive to *H. megidis*, an authentic standard (Sigma-Aldrich, more than 98% pure) was tested in the olfactometer. For this purpose the system was entirely filled with clean moist sand and a 0.2- μ l dose of (*E*)- β -caryophyllene was

injected in the centre of one of the pots, whereas the five remaining pots received no such treatment. Nematodes were released in the middle of the olfactometer and on the next day nematodes were recovered from the six arms. The arm attached to the pot that had received (*E*)- β -caryophyllene contained almost three times as many nematodes as the average control arm (Fig. 2a). Using a much lower dose of three injections with 200 ng of (*E*)- β -caryophyllene in pentane produced very similar results. In an additional experiment, a 10-day-old healthy maize plant (var. Delprim) was placed in each of two opposing pots of the olfactometer and the other four pots contained sand only. One of the pots with a plant was spiked with a 0.2- μ l dose of (*E*)- β -caryophyllene and nematodes were released in the olfactometer centre. On the next day, the arm with the caryophyllene-spiked plant contained on average almost fourfold as many nematodes as the control arms, whereas there was no statistical difference between the plant without (*E*)- β -caryophyllene and the control pots (Fig. 2b). We tested several other synthetic compounds that are commonly released from caterpillar-damaged maize leaves in three choice tests, always including (*E*)- β -caryophyllene as one of the choices (data not shown). These compounds either were not attractive (linalool) or were significantly less attractive than (*E*)- β -caryophyllene ((*Z*)-3-hexenyl acetate, methyl salicylate, (*E*)- β -farnesene, α -humulene and (*E*)-nerolidol) at a 0.2- μ l dose per pot.

Loss of signal in North American maize genotypes

The very limited number of compounds in the volatile blend obtained from WCR-induced roots is in striking contrast to what is emitted from maize leaves in response to caterpillar feeding, a complex mixture of different terpenoids, aromatic compounds and green-leaf volatiles^{4,16,17}. Many of the maize lines that we have screened in the past for caterpillar-induced leaf volatiles do not emit

(*E*)- β -caryophyllene in detectable amounts. This is particularly characteristic for most varieties that originate from North American breeding programmes¹⁸. We tested whether this difference also holds true for the roots by measuring WCR-induced (*E*)- β -caryophyllene in six inbred lines selected from a study on caterpillar-induced leaf emissions¹⁸: three that emitted large amounts (*E*)- β -caryophyllene from their leaves (Du101, F2 and F268) and three lines that released no or very little (*E*)- β -caryophyllene (F584, A654 and F7001). In this experiment we also included the closest wild ancestor of maize, teosinte (*Zea mays parviglumis*^{19,20}), which is known to release relatively large amounts of (*E*)- β -caryophyllene from its leaves in response to caterpillar feeding¹⁷. Ten-day-old plants were subjected to 3 days of WCR feeding, after which (*E*)- β -caryophyllene levels were measured in the roots as above. Teosinte roots were found to release moderate amounts of (*E*)- β -caryophyllene in response to WCR damage (Fig. 3a). The experiment also confirmed a correlation between the levels of (*E*)- β -caryophyllene induced in the leaves and the roots (Fig. 3a): Du101, F2 and F268 emitted considerable amounts of (*E*)- β -caryophyllene from the roots after WCR attack and F584, A654 and F7001 emitted barely detectable amounts. These differences offered an excellent opportunity to test whether (*E*)- β -caryophyllene is a key compound for nematode attraction, because non-emitting varieties should be far less attractive than emitting varieties. This was tested with representative lines of commercial maize for which we had information on caterpillar-induced (*E*)- β -caryophyllene releases¹⁷.

Pactol is a commercial maize variety that releases no detectable amounts of (*E*)- β -caryophyllene from its leaves in response to caterpillar feeding, whereas Graf releases relatively large amounts, significantly more than the variety Delprim, which was used in the first experiments¹⁷. Root extracts from WCR-damaged plants confirmed the presence of (*E*)- β -caryophyllene in Graf and Delprim

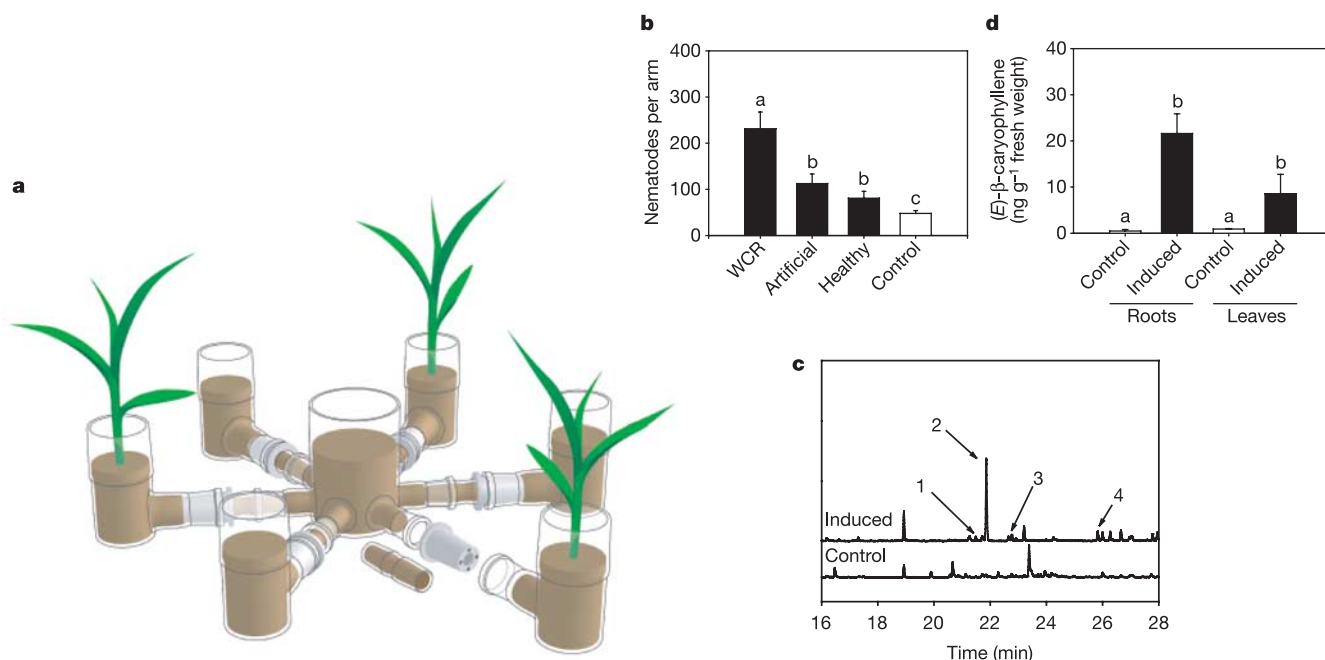


Figure 1 Attraction of entomopathogenic nematodes to a WCR-induced root signal. **a**, Drawing of a newly designed belowground six-arm olfactometer in which nematode attraction was tested. **b**, Choices between plants: the average number of nematodes recovered from olfactometer arms that were connected to pots holding either a maize plant with WCR-damaged roots, mechanically damaged roots or undamaged roots ($n = 12$). For each replicate, the total number of nematodes that went to the three control pots (only moist sand) were summed and divided by three. **c**, Typical

chromatographic traces obtained from the roots of a healthy plant and of a WCR-damaged plant. The labelled peaks are as follows: 1, unknown sesquiterpene; 2, (*E*)- β -caryophyllene; 3, α -humulene; 4, caryophyllene oxide. **d**, Quantification of (*E*)- β -caryophyllene in roots and leaves from healthy and WCR-damaged maize plants ($n = 6$). Letters above bars indicate significant differences. Error bars indicate standard errors.

roots and its absence from Pactol roots (Fig. 3b). Next, individual plants of these three varieties were tested simultaneously in the olfactometer by letting four third-instar WCR larvae feed on their roots for 3 days and then releasing nematodes from the olfactometer centre as before. The numbers of nematodes recovered from the olfactometer arms revealed strong attraction to Graf and Delprim and no attraction to Pactol (Fig. 3c). The importance of (*E*)- β -caryophyllene for this difference in attractiveness was confirmed in a nearly identical experiment with the three varieties, except that on the third day of WCR feeding 0.2 μ l of (*E*)- β -caryophyllene was added to the sand in the pot with the Pactol plant. After this treatment the Pactol plant was as attractive to the nematode as the two other plants (Fig. 3d).

Attractiveness in the field

To verify the importance of (*E*)- β -caryophyllene as an attractant for *H. megidis* under realistic conditions, we conducted two types of field experiment in Hungary, where WCR is already an established pest. For each experiment, six maize plants were planted at an equal distance from each other in circles 1 m in diameter. For the first experiment, three of the plants in each of 33 circles were of the variety Graf, alternated with three plants of the variety Pactol (Fig. 4a). Eight weeks after planting each plant was infested with six second-instar WCR larvae. Seven days after this infestation we released about 10,000 *H. megidis*, three times at 2-day intervals, in the centre of each circle. Larval infection rate by nematodes was determined by collecting the roots with larvae for 15 circles at 3 days after the last nematode release. For the remaining 18 circles, larvae were left to pupate and sleeve cages were placed around the plants at least 1 week before expected adult emergence. In circles with

nematode release, the infection rate for larvae on Graf (43.6% of the recovered larvae) was more than fivefold that for larvae on Pactol (8.3% of the recovered larvae; Fig. 4b). This nematode effect was also evident from a significantly lower emergence of adults from Graf roots (Fig. 4c).

More direct evidence for the importance of (*E*)- β -caryophyllene was obtained with a second experiment with only the Pactol variety planted in the six-plant circles. Again, all plants were infested with six WCR larvae. The soil directly next to three of the plants per circle was spiked on a daily basis with 2 μ l of (*E*)- β -caryophyllene for 5 days (Fig. 5a). One day after the first spiking (7 days after WCR infestation), about 10,000 nematodes were released in the centre of each circle; this was repeated twice at 2-day intervals. We recovered relatively few larvae (18% as opposed to 40% for the Pactol–Graf experiment) from the 12 circles that had been reserved to measure infection rates. This was probably due to poor irrigation of these circles, which could also explain why we did not observe a difference in infection rate between treatments. However, the results from the 24 circles that were left to measure adult emergence showed a significant effect of (*E*)- β -caryophyllene, with a more than twofold decrease in adult emergence for the plants that had been spiked with the signal (Fig. 5b).

The possibility that there could have been a direct effect of (*E*)- β -caryophyllene on the WCR larvae or on the quality of the plant was tested in subsequent laboratory experiments. Equal amounts of (*E*)- β -caryophyllene to those in the field experiments were injected in 15 0.5-litre pots each containing a maize plant and five WCR larvae, whereas 15 other pots each containing a plant and five larvae

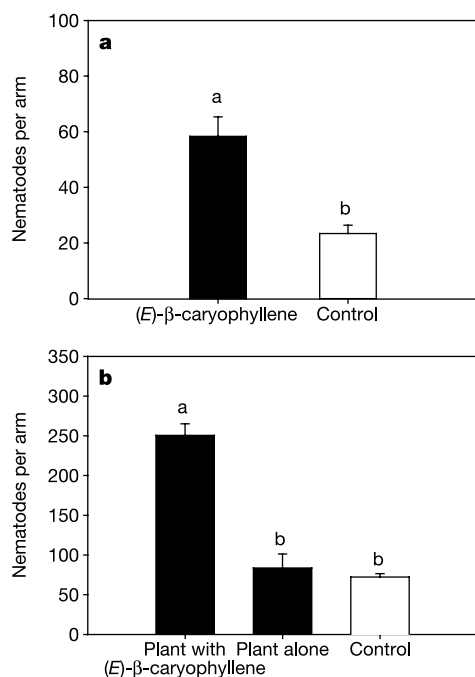


Figure 2 Attraction of *H. megidis* to authentic (*E*)- β -caryophyllene. **a**, Average number of nematodes recovered from olfactometer arms connected to a pot spiked with 0.2 μ l of (*E*)- β -caryophyllene compared with those recovered from arms connected to untreated pots ($n = 12$). **b**, Average number of nematodes recovered from olfactometer arms connected to a pot with a healthy maize plant and spiked with 0.2 μ l of (*E*)- β -caryophyllene, an arm connected to a pot with a healthy plant only, and four control pots with moist sand only ($n = 12$). For each replicate, the results for control pots were summed and divided by the number of control pots. Different letters above bars indicate significant differences. Error bars indicate standard errors.

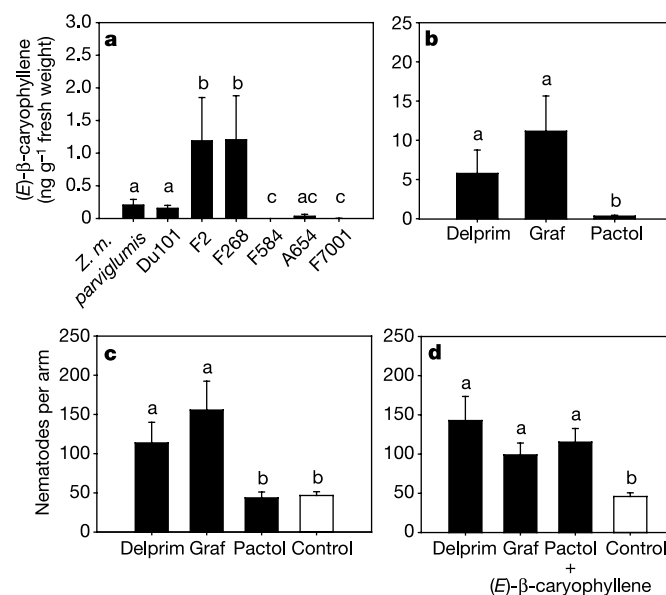


Figure 3 The absence of the (*E*)- β -caryophyllene signal in certain maize genotypes renders these plants unattractive to the nematode. **a**, Average amounts of (*E*)- β -caryophyllene detected from the WCR-damaged roots of *Zea mays parviglumis* (teosinte), of three lines (Du101, F2 and F268) that are known to release (*E*)- β -caryophyllene from their leaves in response to caterpillar damage and of three lines (F584, A654 and F7001) that release no detectable amounts of (*E*)- β -caryophyllene from their leaves¹⁸. **b**, Average amount of (*E*)- β -caryophyllene extracted from WCR-damaged roots of the commercial maize varieties Delprim, Graf and Pactol ($n = 6$). **c**, Average number of nematodes recovered from olfactometer arms connected to pots holding WCR-damaged maize plant of the varieties Delprim, Graf and Pactol ($n = 12$). **d**, Average number of nematodes recovered from olfactometer arms connected to pots holding WCR-damaged maize plant of the varieties Delprim, Graf and Pactol, after the pot with Pactol was spiked with 0.2 μ l of (*E*)- β -caryophyllene ($n = 12$). Statistical differences are indicated with different letters above the bars. Error bars indicate standard errors.

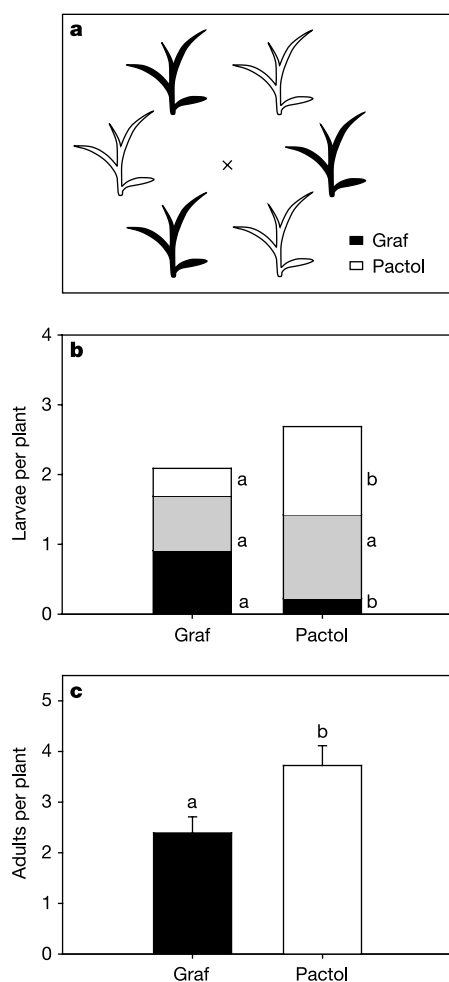


Figure 4 More WCR larvae were infected with nematodes and fewer adults emerged near Graf plants than near Pactol plants. **a**, Design of field circle experiment for which maize plants of the varieties Pactol and Graf were alternated. The cross marks the spot at which nematodes were released. **b**, Mean numbers of larvae per plant that were healthy (white areas), infected by fungi (grey areas) or infected by nematodes (black areas). Statistical differences between the proportions of the three larval types are indicated with different letters. **c**, The mean number of adults that emerged for each plant was significantly different for the two varieties ($P < 0.01$). Error bars indicate standard errors.

served as controls. No difference was found in the total number of adults that emerged from these pots (data not shown), supporting the hypothesis that nematode attraction to (*E*)- β -caryophyllene was responsible for the difference observed in the field.

Suitability of (*E*)- β -caryophyllene as a belowground signal

(*E*)- β -Caryophyllene is a common secondary plant compound that is also emitted from the silk of mature maize plants and has been shown to be weakly attractive to adult WCR females²¹. This sesquiterpene is probably not the only attractant for *H. megidis*, because some degree of nematode attraction was also found to healthy and mechanically damaged plants (Fig. 1b), even though emission of (*E*)- β -caryophyllene from maize leaves and roots has been detected only after herbivory. Indeed, several plant metabolites, including CO₂, are known to be attractants for entomopathogenic nematodes²². Cues that come directly from host larvae might also guide nematodes^{23–26}, but these have been shown to be attractive only over short distances. The overriding importance of (*E*)- β -caryophyllene as a long-range attractant is best indicated by its abundance in the root extracts of the most attractive varieties and

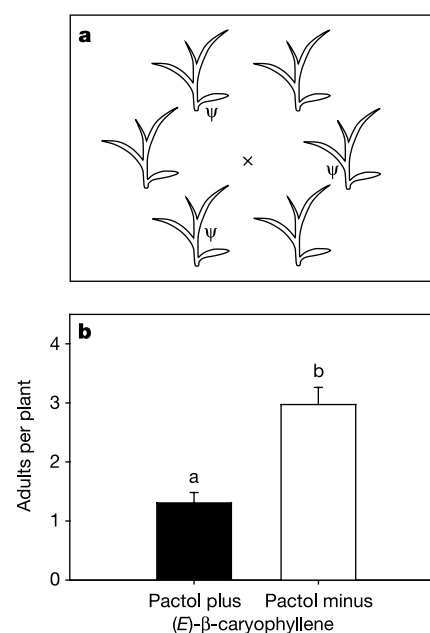


Figure 5 Fewer WCR adults emerged near Pactol plants that were spiked with (*E*)- β -caryophyllene than near Pactol plants that received no (*E*)- β -caryophyllene. **a**, Design of field circle experiment with only plants of the variety Pactol. The ψ signs mark the sites at which five times 2 μ l of (*E*)- β -caryophyllene was injected into the soil; the cross marks the spot at which nematodes were released. **b**, The mean number of adults that emerged near the spiked plants was significantly lower than for the unspiked plants ($P < 0.0001$). Error bars indicate standard errors.

the fact that supplementing sand with (*E*)- β -caryophyllene renders an otherwise unattractive variety highly attractive (Fig. 3d).

To test the ability of (*E*)- β -caryophyllene to diffuse in moist sand, 2 μ g of this sesquiterpene were pipetted into one spot in a sand-filled glass dish. At a distance of 10 cm from this spot a SPME fibre was inserted into a hole in the sand (see Methods). Every half hour the compounds adsorbed on the fibre were desorbed and analysed by GC-MS, starting with the half hour before the addition of (*E*)- β -caryophyllene. (*E*)- β -Caryophyllene travelled rapidly through the sand and was already trapped on the fibre during the first half hour after it had been introduced to the sand. The amount trapped increased steadily for 2 h, after which it decreased sharply (Fig. 6a). A similar experiment in a sand-filled olfactometer, with an arm modified to permit the introduction of a SPME fibre, revealed the presence of (*E*)- β -caryophyllene in the centre part of an arm 2 h after injecting 0.2 μ l into a pot connected to that arm (not shown). To determine whether the rapid decrease in (*E*)- β -caryophyllene detection was due to evaporation from the sand, an additional experiment was performed by which a drop containing 1 μ g of (*E*)- β -caryophyllene was placed on the bottom of a beaker, which was immediately covered by 5 cm of moist sand. The beaker was placed in a closed-loop volatile-collection system where the headspace above the sand was continuously sampled at intervals of 30 min. A very similar time course of (*E*)- β -caryophyllene diffusion was obtained, with the first detection after 30 min and a peak after 2 h (Fig. 6b), indicating rapid evaporation. Recovery was more than 90%, which implies that the degradation of (*E*)- β -caryophyllene or its immobilization to sand particles is not significant under these conditions. The rapid diffusion of (*E*)- β -caryophyllene in moist sand and its chemical stability seem to make it exceptionally suitable as a belowground signal. In the olfactometer assays described above, the nematodes were released 25 cm from the treatment pots. Therefore, after detecting the signal they move a distance of more than 250 times their body length within a day.

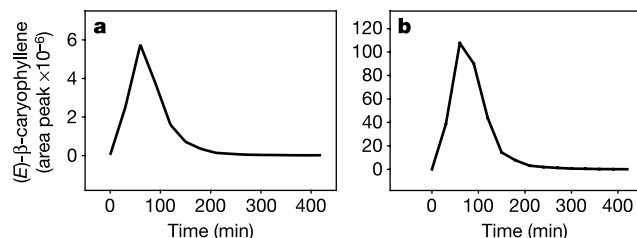


Figure 6 (*E*)- β -Caryophyllene diffuses readily through sand and then evaporates rapidly without breakdown or irreversible adsorption. **a**, Detection of authentic (*E*)- β -caryophyllene with a SPME fibre in moist sand at 10 cm from a release point, every half hour after release. **b**, Detection of (*E*)- β -caryophyllene in the headspace above a beaker containing 5 cm of moist sand after (*E*)- β -caryophyllene had been placed at the bottom of the beaker.

Discussion

The failure of most North American maize lines to release (*E*)- β -caryophyllene suggests that the ability to produce this compound has been lost during breeding. Indeed, the closest wild ancestor of maize, *Zea mays* ssp. *parviglumis*^{19,20}, was also found to release (*E*)- β -caryophyllene from its roots in response to WCR damage (Fig. 3a). The loss of direct defences to herbivores during plant domestication has been amply documented²⁷. However, to our knowledge this is the first example of the loss of a signal involved in indirect defence.

WCR has already caused large economic losses to maize in Central Europe. Since 2003 it has been detected in almost all European countries south of Scandinavia, and will inevitably become a major threat to maize cultivation throughout Europe²⁸. Effective, ecologically sound control methods are needed. Entomopathogenic nematodes could be an option^{14,29–31}, but they have not yet been employed with sufficient efficacy. The results of this study lead us to speculate that the absence of an attractive signal in many American maize lines could explain why attempts to control WCR with nematodes have yielded only mixed results on the North American continent^{32,33}. Reintroduction of this signal in newly developed maize varieties might aid in effective control of this voracious pest.

This first identification of an inducible belowground plant signal that attracts enemies of root-feeding herbivores underscores the breadth and sophistication of indirect plant defences. With a growing interest in belowground plant-mediated interactions and their effects on various trophic levels^{34,35} our results should prompt new studies into the evolutionary history and ecological consequences of multitrophic-level interactions and should lead to the exploitation of the signal for crop protection. □

Methods

Olfactometer assays

The attraction of nematodes to plant-produced substances was tested in a belowground olfactometer consisting of a central glass chamber (8 cm in diameter, 11 cm deep) with six equally distributed side arms at 0.5 cm height with a female (24 mm diameter $\times$ 29 mm long) connector (Fig. 1a). These arms connected the central chamber with six glass pots (5 cm in diameter, 11 cm deep) in which plants or other sources of attractants could be placed. Each pot also had a female connector (29/32) at 0.5 cm height. The connecting arms consisted of two detachable parts; one was a glass tube with ground-glass connectors (male, 24/29) on both sides, and the second part, a Teflon connector (24/29 to 29/32) was used to attach the glass tube to the odour source pot. The custom-made Teflon connectors (Analytical Research Systems) contained an ultra-fine metal screen (2,300 mesh; Small Parts Inc.) preventing the nematodes from reaching the odour source pots (Fig. 1a). For each experiment, the entire system was filled with sterilized white sand (Migros) to about 5 cm from the rim of the pots. Nematodes were released in a drop of water in the centre of the central pot. One day after nematode release, the olfactometer was disassembled and the sand in each detachable glass tube was placed on a separate cotton filter disk 19 cm in diameter (Hoeschele GmbH). The disk with the sand was placed in a Bearmann extractor^{36,37}, and nematodes in the collection tube were counted on the next day.

Statistical differences in choices made by the nematodes were determined with log-linear models on the basis of the assumption that the nematodes would disperse equally among the arms in the absence of any attraction. The models were adapted to account for possible overdispersion due to directional biases³⁸.

Root analyses

For the analysis of volatile terpenes, roots of WCR-damaged and undamaged maize plants were washed with water and frozen in liquid nitrogen; they were then pulverized in a mortar and 0.4 g of root powder was placed in a glass vial with a septum in the lid. A 100- μ m polydimethylsiloxane (PDMS) SPME (Supelco) fibre was inserted through the septum and exposed for 60 min at 40 °C. The compounds adsorbed on the fibre were analysed by GC–MS with an Agilent 6890 Series GC system G1530A coupled to a quadrupole-type mass-selective detector (Agilent 5973; transfer line 230 °C, source 230 °C, ionization potential 70 eV). The fibre was inserted manually into the injector port (230 °C) and desorbed and chromatographed on an apolar column (DB5-MS, 30 m, 0.25 mm internal diameter, 0.25 μ m film thickness; J & W Scientific). Helium at a constant pressure of 18.55 lb in⁻² (127.9 kPa) was used for carrier gas flow. After fibre insertion, the column temperature was maintained at 50 °C for 3 min and then increased to 180 °C at 5 °C min⁻¹ followed by a final stage of 3 min at 250 °C. Approximate quantification was performed with an external standard by performing analyses on 0.4 g of powdered root tissue from maize line B73 (which produces only traces of (*E*)- β -caryophyllene)³⁹ spiked with known amounts (4.5, 9.0, 45 and 90 ng) of this compound.

The (*E*)- β -caryophyllene in the roots was provisionally identified as the (–)-enantiomer by chromatography on a chiral column using published procedures for the separation of the two enantiomers⁴⁰. However, the lack of a standard for the (+)-enantiomer prevented final confirmation.

Field experiments

Field experiments were conducted at the Plant Health Station in Hodmészövárshely, in southern Hungary (46° 15.554' N, 20° 09.743' E), from April to October 2004. Six plants were grown from seed in 1-m-diameter circles and with a 1-m distance between circles.

Two types of circle were formed: one contained the two varieties *Zea mays* var. Pactol (Syngenta) and *Z. mays* var. Graf (Landi) and the other contained only the Pactol variety. Eight weeks after planting, each plant was infested with six WCR larvae by digging out 5 cm of soil near the base of the plant and dropping the larvae with some potting soil into the hole. The larvae came from a laboratory colony that had been established with field-collected adults the year before. At 7, 9 and 11 days after infestation, about 10,000 *H. megidis* nematodes were released in the centre of the treatment circles at a depth of about 10 cm. Additionally, half of the plants in the circles with only the Pactol variety were spiked daily with 2 μ l of (*E*)- β -caryophyllene (more than 98% pure; Sigma-Aldrich) for 5 days, starting on the sixth day after infestation with WCR larvae (1 day before nematode release).

Two measurements were taken to determine the effect of the treatments on nematode effectiveness. For 15 of the Pactol–Graf circles and 12 of the Pactol–Pactol circles, the aerial part of the plants was removed and with a 1-litre core sample the roots and soil around it were collected. Larvae were extracted by crumbling the soil over a black plastic sheet and dissecting the roots. Each recovered larva was placed on a moist filter paper in a plastic Petri dish (5 cm in diameter, 2 cm deep) and stored at 17 °C for 1 month. They were checked weekly under a microscope for nematode infection, characterized by red pigmentation resulting from symbiotic bacteria, and for nematode emergence. Infections by other pathogens were also noted.

In addition, adult emergence was measured in another 18 Pactol–Graf circles and 24 Pactol–Pactol circles. For this, cylindrical sleeve cages (30 cm $\times$ 70 cm, MegaView Science Education Services Co. Ltd) were fixed on plastic cylinders 20 cm in diameter and 25 cm deep that were placed about 10 cm in the soil around each plant. The upper part of each sleeve was tightly attached around the stem of the plant to prevent adults from escaping. Once a week, from the beginning of July until the end of August, adults in the emergence cages were counted and collected until no more adults were found. The same log-linear models as employed for the olfactometer data were used to determine differences between treatments³⁸.

Diffusion measurements

A glass dish (15 cm in diameter, 8 cm deep) was filled with a 5-cm layer of moist (10% water) sand. With a micropipette, 2 μ g of authentic (*E*)- β -caryophyllene (98% pure; Aldrich) in 10 μ l of pentane was placed 3 cm deep in the sand at 2 cm from the dish side, immediately after which the hole was covered. At a distance of 10 cm from this spot, a hole 2 mm wide and 3 cm deep was made with a metal rod and a 100- μ m PDMS SPME fibre was placed in the hole. Every 25 min the compounds adsorbed on the fibre were analysed by GC–MS essentially as described above, except that the mass selective detector was operated in the selective ion mode, scanning only for the characteristic ions at molecular masses 204, 133 and 93. After the 5-min desorption period, the fibre was placed back in the hole in the sand for a further 25-min collection. The first collection started 30 min before the (*E*)- β -caryophyllene sample was added to the sand, and the last collection was 7 h later.

To measure the time course of evaporation from sand, a 10- μ l drop of dichloromethane containing 1 μ g of (*E*)- β -caryophyllene was placed on the bottom of a 25-mm diameter glass beaker and was immediately covered by a 5-cm layer of 30 ml moist sand. The beaker was placed in a closed-loop volatile collection system consisting of a 1-litre desiccator in which the headspace above the sand was continuously collected by pulling air through a 75-mg activated charcoal filter at a rate of 21 min⁻¹. The filter was extracted with dichloromethane at intervals of 30-min and the eluate was analysed by GC–MS as described above.

Received 15 November 2004; accepted 8 February 2005; doi:10.1038/nature03451.

- Karban, R. & Baldwin, I. *Induced Responses to Herbivory* (Univ. Press Chicago, Chicago, 1997).
- Agrawal, A. A., Tuzun, S. & Bent, E. *Induced Plant Defenses Against Pathogens and Herbivores* (APS Press, St Paul, Minnesota, 1999).
- Dicke, M. & Sabelis, M. W. How plants obtain predatory mites as bodyguards. *Neth. J. Zool.* **38**, 148–165 (1988).
- Turlings, T. C. J., Tumlinson, J. H. & Lewis, W. J. Exploitation of herbivore-induced plant odors by host-seeking parasitic wasps. *Science* **250**, 1251–1253 (1990).
- De Moraes, C. M., Lewis, W. J., Paré, P. W., Alborn, H. T. & Tumlinson, J. H. Herbivore-infested plants selectively attract parasitoids. *Nature* **393**, 570–573 (1998).
- Thaler, J. S. Jasmonate-inducible plant defences cause increased parasitism of herbivores. *Nature* **399**, 686–688 (1999).
- Kessler, A. & Baldwin, I. T. Defensive function of herbivore-induced plant volatile emissions in nature. *Science* **291**, 2141–2144 (2001).
- Khan, Z. R. *et al.* Intercropping increases parasitism of pests. *Nature* **388**, 631–632 (1997).
- Boff, M. I. C., Zoon, F. C. & Smits, P. H. Orientation of *Heterorhabditis megidis* to insect hosts and plant roots in a Y-tube sand olfactometer. *Entomol. Exp. Appl.* **98**, 329–337 (2001).
- van Tol, R. W. H. M. *et al.* Plants protect their roots by alerting the enemies of grubs. *Ecol. Lett.* **4**, 292–294 (2001).
- Osteen, C. in *Agricultural Resources and Environmental Indicators, 2003* (ed. Heimlich, R.) 1–44 (Economic Research Service, Washington DC, 2003).
- Krysan, J. L. & Miller, T. A. *Methods for the Study of the Pest Diabrotica* (Springer, New York, 1986).
- Kuhlmann, U. & van der Burgt, W. A. C. M. Possibilities for biological control of the western corn rootworm, *Diabrotica virgifera virgifera* LeConte, in Central Europe. *BioControl* **19**, 59N–68N (1998).
- Kaya, H. K. & Gaugler, R. Entomopathogenic nematodes. *Annu. Rev. Entomol.* **38**, 181–206 (1993).
- Gaugler, R. Ecological considerations in the biological-control of soil-inhabiting insects with entomopathogenic nematodes. *Agric. Ecosyst. Environ.* **24**, 351–360 (1988).
- Turlings, T. C. J., Lengwiler, U. B., Bernasconi, M. L. & Wechsler, D. Timing of induced volatile emissions in maize seedlings. *Planta* **207**, 146–152 (1998).
- Gouinguene, S., Degen, T. & Turlings, T. C. J. Variability in herbivore-induced odour emissions among maize cultivars and their wild ancestors (teosinte). *Chemoecology* **11**, 9–16 (2001).
- Degen, T., Dillmann, C., Marion-Poll, F. & Turlings, T. C. J. High genetic variability in herbivore-induced volatile emission within a broad range of maize inbred lines. *Plant Physiol.* **135**, 1928–1938 (2004).
- Fedoroff, N. V. Prehistoric GM corn. *Science* **302**, 1158–1159 (2003).
- Jaenicke-Despres, V. *et al.* Early allelic selection in maize as revealed by ancient DNA. *Science* **302**, 1206–1208 (2003).
- Hammack, L. Single and blended maize volatiles as attractants for diabrotic corn rootworm beetles. *J. Chem. Ecol.* **27**, 1373–1390 (2001).
- O'Halloran, D. M. & Burnell, A. M. An investigation of chemotaxis in the insect parasitic nematode *Heterorhabditis bacteriophora*. *Parasitology* **127**, 375–385 (2003).
- Grewal, P. S., Lewis, E. E., Gaugler, R. & Campbell, J. F. Host finding behavior as a predictor of foraging strategy in entomopathogenic nematodes. *Parasitology* **108**, 207–215 (1994).
- Grewal, P. S., Lewis, E. E. & Gaugler, R. Response of infective stage parasites (Nematoda: Steinernematidae) to volatile cues from infected hosts. *J. Chem. Ecol.* **23**, 503–515 (1997).
- Lewis, E. E., Gaugler, R. & Harrison, R. Response of cruiser and ambusher entomopathogenic nematodes (Steinernematidae) to host volatile cues. *Can. J. Zool.* **71**, 765–769 (1993).
- Lewis, E. E., Grewal, P. S. & Gaugler, R. Hierarchical order of host cues in parasite foraging strategies. *Parasitology* **110**, 207–213 (1995).
- Sotelo, A. in *Functionality of Food Phytochemicals* (eds Jones, T. & Romeo, J.) 89–111 (Plenum, New York, 1997).
- Baufeld, P. & Enzian, S. in *Western Corn Rootworm: Ecology and Management* (eds Vidal, S., Kuhlmann, U. & Edwards, C. R.) 285–302 (CABI, Wallingford, 2005).
- Journey, A. M. & Ostlie, K. R. Biological control of the western corn rootworm (Coleoptera: Chrysomelidae) using the entomopathogenic nematode, *Steinernema carpocapsae*. *Environ. Entomol.* **29**, 822–831 (2000).
- Levine, E. & Oloumisadeghi, H. Management of diabrotic rootworms in corn. *Annu. Rev. Entomol.* **36**, 229–255 (1991).
- Poinar, G. O., Evans, J. S. & Schuster, E. Field test of the entomogenous nematode, *Neoplectana carpocapsae*, for control of corn-rootworm larvae (*Diabrotica* sp., Coleoptera). *Protection Ecol.* **5**, 337–342 (1983).
- Ellsbury, M. M., Jackson, J. J., Woodson, W. D., Beck, D. L. & Stange, K. A. Efficacy, application distribution, and concentration by stemflow of *Steinernema carpocapsae* (Rhabditida: Steinernematidae) suspensions applied with a lateral-move irrigation system for corn rootworm (Coleoptera: Chrysomelidae) control in maize. *J. Econ. Entomol.* **89**, 74–81 (1996).
- Jackson, J. J. Field performance of entomopathogenic nematodes for suppression of western corn rootworm (Coleoptera: Chrysomelidae). *J. Econ. Entomol.* **89**, 366–372 (1996).
- van der Putten, W. H., Vet, L. E. M., Harvey, J. A. & Waeckers, F. L. Linking above- and belowground multitrophic interactions. *Trends Ecol. Evol.* **16**, 547–554 (2001).
- Wardle, D. A. *et al.* Ecological linkages between aboveground and belowground biota. *Science* **304**, 1629–1633 (2004).
- Hass, B., Griffin, C. T. & Downes, M. J. Persistence of *Heterorhabditis* infective juveniles in soil: Comparison of extraction and infectivity measurements. *J. Nematol.* **31**, 508–516 (1999).
- Curran, J. Influence of application method and pest population-size on the field efficacy of entomopathogenic nematodes. *J. Nematol.* **24**, 631–636 (1992).
- Turlings, T. C. J., Davison, A. C. & Tamo, C. A six-arm olfactometer permitting simultaneous observation of insect attraction and odour trapping. *Physiol. Entomol.* **29**, 45–55 (2004).
- Köllner, T. G., Schnee, C., Gershenzon, J. & Degenhardt, J. The sesquiterpene hydrocarbons of maize (*Zea mays*) form five groups with distinct developmental and organ-specific distribution. *Phytochemistry* **65**, 1895–1902 (2004).
- König, W. A., Bulow, N. & Saritas, Y. Identification of sesquiterpene hydrocarbons by gas phase analytical methods. *Flav. Fragr. J.* **14**, 367–378 (1999).

Acknowledgements We thank the members of the group of M. Rahier for their continuous support; and M.-E. Farine for technical assistance; R. Gaugler, J. Tumlinson and R. van Tol for useful feedback; and V. Larraz for rearing the *Diabrotica* larvae for laboratory experiments. Field sites were provided by the Plant Health Service in Hodmészövásárhely, Hungary. Nematodes were provided, free of charge, by Andermatt Biocontrol AG. The work was supported by the Swiss National Centre of Competence in Research 'Plant Survival', the Swiss National Science Foundation, the German National Science Foundation and the Max Planck Society.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.C.J.T. (ted.turlings@unine.ch).

Rapid growth of black holes in massive star-forming galaxies

D. M. Alexander¹, I. Smail², F. E. Bauer¹, S. C. Chapman³, A. W. Blain³, W. N. Brandt⁴ & R. J. Ivison^{5,6}

¹Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

²Institute for Computational Cosmology, University of Durham, South Road, Durham DH1 3LE, UK

³California Institute of Technology, Pasadena, California 91125, USA

⁴Department of Astronomy and Astrophysics, Pennsylvania State University, 525 Davey Laboratory, University Park, Pennsylvania 16802, USA

⁵Astronomy Technology Centre, Royal Observatory, and ⁶Institute for Astronomy, University of Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK

The tight relationship between the masses of black holes and galaxy spheroids in nearby galaxies¹ implies a causal connection between the growth of these two components. Optically luminous quasars host the most prodigious accreting black holes in the Universe, and can account for ≥ 30 per cent of the total cosmological black-hole growth^{2,3}. As typical quasars are not, however, undergoing intense star formation and already host massive black holes ($>10^8 M_\odot$, where $M_\odot$ is the solar mass)^{4,5}, there must have been an earlier pre-quasar phase when these black holes grew (mass range $\sim(10^6\text{--}10^8)M_\odot$). The likely signature of this earlier stage is simultaneous black-hole growth and star formation in distant (redshift $z > 1$; >8 billion light years away) luminous galaxies. Here we report ultra-deep X-ray observations of distant star-forming galaxies that are bright at submillimetre wavelengths. We find that the black holes in these galaxies are growing almost continuously throughout periods of intense star formation. This activity appears to be more tightly associated with these galaxies than any other coeval galaxy populations. We show that the black-hole growth from these galaxies is consistent with that expected for the pre-quasar phase.

The most intense sites of star formation at high redshift are associated with submillimetre galaxies (SMGs)^{6–9}. SMGs are amongst the most bolometrically luminous galaxies in the Universe (bolometric luminosity $L_{\text{BOL}} \approx 10^{13} L_\odot$, where $L_\odot$ is the solar luminosity) and the majority of the population is coeval with the peak in quasar activity (that is, $z \approx 1.5\text{--}3$)^{7,9}. The apparent association of SMGs with some quasars and the similarity in the comoving space densities of SMGs and optically luminous (B-band magnitude $M_B < -24$) quasars (when corrected for probable source lifetimes) provides direct evidence for an evolutionary connection between SMGs and quasars^{5,10,11}. However, in contrast to quasars, the bolometric output of SMGs appears to be dominated by powerful star-formation activity and any black-hole accretion—that is, active galactic nuclei (AGN) activity—is comparatively weak^{12,13}. Given the large available molecular gas supply (typically $\sim 3 \times 10^{10} M_\odot$)¹⁴, SMGs can fuel this luminous star-formation activity for $\sim 10^8$ yr (refs 14, 15). Various items of complementary observational support have also shown that SMGs are massive galaxies ($\sim 10^{11} M_\odot$), suggesting that they will become $\geq M^*$ spheroid-dominated galaxies in the local Universe^{14,16,17} (M^* is the mass of a typical massive galaxy).

AGN activity has been identified in many SMGs^{6,12}. However, the difficulty in obtaining high-quality optical spectra and reliable source redshifts has hindered a complete census of AGN activity in SMGs. We have initiated a project investigating the properties of AGNs in SMGs using deep optical spectroscopic data (obtained with the 10-m Keck telescope)⁹ and ultra-deep X-ray observations (Chandra Deep Field-North; CDF-N)¹⁸. As X-ray emission appears to be a ubiquitous property of AGNs that is comparatively impervious to obscuration, the combination of these deep data sets

provides a detailed census of AGN activity in SMGs. Here we focus on the role of black-hole accretion in SMGs and the growth of black holes in massive galaxies. The assumed cosmology is $H_0 = 65 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_M = 1/3$, and $\Omega_\Lambda = 2/3$.

The parent SMG sample comprises 20 submillimetre-detected (with the SCUBA camera on the James Clerk Maxwell Telescope) sources in the CDF-N region whose positions have been reliably located from their radio emission and which could then be spectroscopically targeted⁹. These sources have overall properties consistent with the general SMG population (850- μm flux $S_{850\mu\text{m}} \approx 4\text{--}12 \text{ mJy}$ and $z = 0.56\text{--}2.91$, with the majority at $z > 2$)^{6–9}. Seventeen ($\sim 85\%$) of the 20 SMGs are detected at X-ray energies. A detailed investigation of the nature of the X-ray emission has revealed AGN activity in at least 15 ($\sim 75\%$) SMGs¹⁹; the properties of these sources are given in Table 1. The X-ray properties of the other 5 ($\sim 25\%$) SMGs are consistent with those expected from luminous star-formation activity, but we note that at least one shows evidence for AGN activity at near-infrared wavelengths¹⁶.

Two different observing modes were used in constructing the SMG sample: 14 of the radio sources were specifically targeted with SCUBA observations and 6 of the radio sources were detected in blank-field SCUBA maps. Although targeting known radio sources with SCUBA observations could potentially bias our AGN fraction owing to contributions to the radio emission from AGN activity, a statistical analysis does not reveal a strong AGN bias in our sample ($P = 1.0$, two-sided Fisher's exact test); see Table 1 for the observing modes of the X-ray classified AGNs. However, owing to the large positional uncertainties of SCUBA sources ($\sim 6\text{--}7$ arcsec), redshifts could not be obtained for the $\sim 35\text{--}50\%$ of the radio-undetected SMG population, some of which might host AGN activity^{6,9}. Making the conservative assumption that none of the radio-undetected SMGs hosts AGN activity, our X-ray data alone suggest an AGN fraction in the SMG population of $> 38^{+12}_{-10}\%$. We note that a similar fraction of comparably luminous galaxies in the local Universe host AGN activity²⁰, although their space density is approximately three orders of magnitude smaller than that of SMGs.

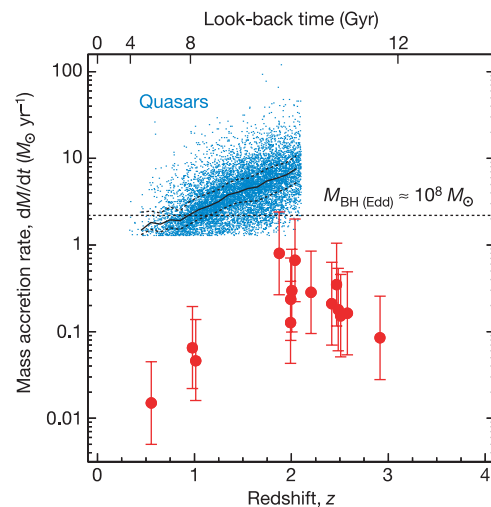


Figure 1 Black-hole mass accretion rates. SMGs (filled red circles) and a comparison sample of $M_B < -24$ quasars (small blue dots) are shown⁴. The error bars represent the estimated uncertainty in the X-ray to bolometric correction for AGN activity in SMGs; see Table 1. The solid and dashed curves indicate the median and interquartile ranges for the comparison quasar sample. The dotted line indicates the approximate Eddington-limited mass accretion rate for an $\sim 10^8 M_\odot$ black hole ($M_{\text{BH (Edd)}}$). The SMGs have mass accretion rates that are approximately an order of magnitude lower than those of coeval quasars⁴, suggesting that their black holes are smaller (typically $\sim (10^7\text{--}10^8)M_\odot$); see text for additional black-hole mass constraints.

Table 1 The properties of the X-ray classified AGNs

SMG name	$S_{850\mu\text{m}}$ (mJy)	z	$\log(L_X \text{ (erg s}^{-1}\text{)})$	$\log(L_{\text{BOL}} \text{ (} L_\odot\text{)})$	$dM/dt \text{ (} M_\odot \text{ yr}^{-1}\text{)}$	$M_{\text{BH (Edd)}} \text{ (} 10^7 M_\odot\text{)}$	X-ray obsc.?	Obs. mode?
123549.4 + 621536	8.3 ± 2.5	2.20	44.0	12.9	$0.28 (+0.57, -0.19)$	$1.3 (+2.6, -0.9)$	Y	T
123555.1 + 620901	5.4 ± 1.9	1.88	44.4	13.2	$0.80 (+1.60, -0.53)$	$3.6 (+7.3, -2.4)$	Y	T
123606.7 + 621550	4.4 ± 1.4	2.42	43.8	12.6	$0.21 (+0.42, -0.14)$	$1.0 (+1.9, -0.6)$	N	M
123606.8 + 621021	11.6 ± 3.5	2.51	43.7	13.1	$0.15 (+0.31, -0.10)$	$0.7 (+1.4, -0.5)$	Y	T
123616.1 + 621513	5.8 ± 1.1	2.58	43.7	13.0	$0.16 (+0.33, -0.11)$	$0.7 (+1.5, -0.5)$	Y	T
123622.6 + 621629	7.7 ± 1.3	2.47	44.0	13.1	$0.35 (+0.70, -0.23)$	$1.6 (+3.2, -1.1)$	Y	T
123629.1 + 621045	5.0 ± 1.3	1.01	43.2	12.1	$0.05 (+0.09, -0.03)$	$0.2 (+0.4, -0.1)$	Y	T
123632.6 + 620800	5.5 ± 1.3	1.99	43.9	12.9	$0.24 (+0.47, -0.16)$	$1.1 (+2.2, -0.7)$	Y	M
123635.5 + 621424	5.5 ± 1.4	2.01	44.0	12.9	$0.30 (+0.60, -0.20)$	$1.4 (+2.7, -0.9)$	Y	T
123636.7 + 621156	7.0 ± 2.1	0.56	42.7	11.1	$0.02 (+0.03, -0.01)$	$0.1 (+0.1, -0.1)$	N	M
123707.2 + 621408	4.7 ± 1.5	2.48	43.8	12.9	$0.18 (+0.36, -0.12)$	$0.8 (+1.6, -0.5)$	Y	T
123711.9 + 621325	4.2 ± 1.4	1.99	43.6	12.7	$0.13 (+0.25, -0.08)$	$0.6 (+1.2, -0.4)$	Y	T
123712.0 + 621212	8.0 ± 1.8	2.91	43.4	12.7	$0.09 (+0.17, -0.06)$	$0.4 (+0.8, -0.3)$	Y	M
123716.0 + 620323	5.3 ± 1.7	2.04	44.3	13.0	$0.67 (+1.33, -0.44)$	$3.0 (+6.1, -2.0)$	Y	T
123721.8 + 621035	12.0 ± 3.9	0.98	43.7	11.7	$0.16 (+0.31, -0.10)$	$0.7 (+1.4, -0.5)$	Y	M

SMG name, submillimetre flux density ($S_{850\mu\text{m}}$), redshift (z) and total bolometric luminosity (L_{BOL}) are taken from ref. 9. Unobscured rest-frame 0.5–8.0 keV luminosity (L_X), and the presence of X-ray obscuration ('X-ray obsc.?', $N_{\text{H}} > 10^{22} \text{ cm}^{-2}$) are taken from ref. 19. The observation mode ('Obs. mode?') refers to whether the source was specifically targeted with a SCUBA observation ('T') or whether the source was detected in a blank-field SCUBA map ('M'): 10 of the AGNs were identified in targeted SCUBA observations and 5 of the AGNs were identified in blank-field SCUBA maps. The mass accretion rate (dM/dt) is estimated by converting the X-ray luminosity to the AGN bolometric luminosity, under the assumption that the X-ray luminosity accounts for $6^{+12}_{-4}\%$ of the bolometric luminosity of the AGN (that is, the range in conversion factors found in ref. 29). Although our assumed bolometric conversion factors were originally derived from unobscured AGNs, whereas the majority of the AGNs in our sample are obscured, the important issue is the relationship between the intrinsic X-ray emission and the ultraviolet-optical accretion disk emission, which is unlikely to be obscuration dependent. We also note that our bolometric conversion factors agree with those estimated for obscured AGNs with similar luminosities (and therefore similar mass accretion rates) to the sources studied here³⁰. When converting from the bolometric luminosity to the accreted mass we assumed the canonical efficiency of 10% ($\epsilon = 0.1$). The black-hole mass ($M_{\text{BH (Edd)}}$) is calculated from the mass accretion rate under the assumption that the accretion is Eddington limited; see text for justification. The black-hole masses will be higher if the accretion is sub-Eddington; see text for additional black-hole mass constraints.

The large AGN fraction in the SMG population indicates that their black holes are growing almost continuously throughout the intense star-formation phase of these galaxies—that is, assuming that all SMGs are growing their black holes, the AGN duty cycle is $> 38^{+12}_{-10}\%$. This suggests that the black holes and galaxy spheroids are growing concordantly in SMGs¹. Such a close association between AGN and star-formation activity is not seen in the coeval field galaxy population or other coeval star-forming galaxy populations ($\sim 3\text{--}15\%$ AGN fraction)^{21,22}. The almost continuous black-hole growth in SMGs suggests that there is an abundance of available fuel, hinting that the accretion may be occurring efficiently (that is, at or close to the Eddington limit), as predicted by theoretical studies of the growth of massive black holes^{23–25}. The majority ($\sim 85\%$) of the AGNs are obscured (see Table 1), as also predicted for black holes undergoing efficient growth²³.

The unobscured X-ray luminosities of the AGNs (X-ray luminosity $L_X \approx 10^{43}\text{--}10^{44} \text{ erg s}^{-1}$) suggest that the mass accretion rates are modest ($\lesssim 1 M_\odot \text{ yr}^{-1}$); see Table 1 and Fig. 1. By comparison to coeval quasars, the mass accretion rates in SMGs are approximately an order of magnitude lower, also suggesting modest black-hole masses (that is, $\sim (10^7\text{--}10^8) M_\odot$; see Table 1 and Fig. 1). Although these black-hole mass estimates are uncertain, the narrowness of the broad emission lines (BELs; typical full-width at half-maximum velocities of $\sim 2,500 \text{ km s}^{-1}$)¹⁶, when they are visible, also suggest black-hole masses of $< 10^8 M_\odot$ (ref. 4). As well-studied AGNs with narrow BELs are found to be accreting at (or even beyond) the Eddington limit⁴, this further supports the idea that the mass accretion rate in SMGs is approximately Eddington limited. Under the assumption of Eddington-limited accretion, the black holes in the SMGs will grow by up to an order of magnitude over their $\sim 10^8 \text{ yr}$ star-formation lifetime. These overall properties are consistent with those expected for massive galaxies undergoing rapid black-hole growth.

The total black-hole growth from SMGs can be calculated by integrating the average accretion density over the SMG redshift range (Fig. 2). Over the redshift interval $z = 1.8\text{--}3.0$ (corresponding to 80% of the SMGs studied here), the black-hole density produced by SMGs is $\sim 6^{+11}_{-4} \times 10^3 M_\odot \text{ Mpc}^{-3}$. To put this quantity into context, we need to compare it to the black-hole growth from coeval quasar activity. Considering only quasars in the luminosity range of $M_B = -24$ to -27 (the majority of which reside in $\geq M^*$ spheroid-dominated galaxies²⁶ and have the properties consistent with being the progeny of SMGs⁵), the total SMG black-hole density

is $\sim 13^{+27}_{-9}\%$ of the black-hole growth from quasar activity². These constraints should be considered lower limits, as further pre-quasar growth at $z = 1.8\text{--}3.0$ may be produced by radio-undetected SMGs and galaxies with fainter submillimetre fluxes, if they host AGN activity. However, these results are consistent with, for example, SMGs undergoing an intense black-hole growth phase, where the black hole is grown from $\sim 10^7 M_\odot$ to $\sim 10^8 M_\odot$, before a high-accretion-rate quasar phase, where the black hole is grown from $\sim 10^8 M_\odot$ to $\sim 8 \times 10^8 M_\odot$. In this scenario, although SMGs do not produce a large fraction of the cosmological black-hole density, they are responsible for the crucial pre-quasar phase when the black holes in massive galaxies are rapidly grown. Indeed, the total black-hole growth from SMGs could only exceed that from quasar activity if the quasar lifetime is insufficient to double the mass of the black hole (that is, $< 3 \times 10^7 \text{ yr}$ assuming Eddington-limited accretion).

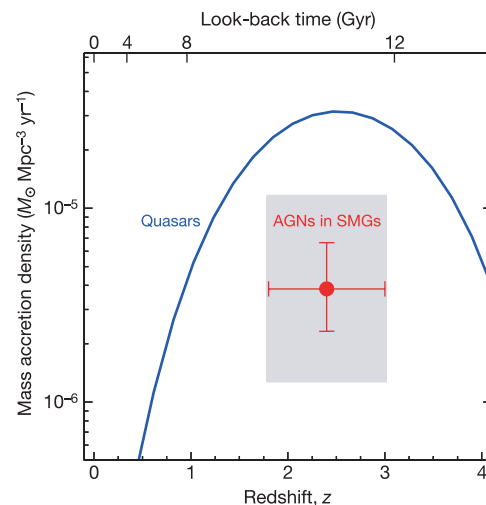


Figure 2 Cosmological black-hole accretion density. SMGs (filled red circle), and quasars with $M_B = -24$ to -27 (blue solid line)², are shown. Error bars in the x -axis direction indicate the redshift range, and error bars in the y -axis direction represent 1σ Poisson counting uncertainties. The shaded region indicates the uncertainty in the mass accretion rate conversion for the SMGs; see Table 1. The mass accretion density from SMGs is $\sim 13^{+27}_{-9}\%$ of that produced by coeval $M_B = -24$ to -27 quasar activity over the redshift interval $z = 1.8\text{--}3.0$.

Overall, this picture is in good agreement with direct theoretical predictions of the black-hole growth of SMGs and quasars^{24,25}.

What was the catalyst for the rapid black-hole and stellar growth seen in SMGs? Rest-frame ultraviolet images taken by the Hubble Space Telescope have shown that a considerably larger fraction of SMGs are in major mergers (that is, the merging of two galaxies of comparable masses) than has been found in the coeval galaxy population^{17,27}. Hydrodynamical simulations have shown that major mergers can efficiently transport material towards the central regions of galaxies, providing an effective mechanism to trigger nuclear star formation and fuel the black hole²⁸. The result of these major mergers is thought to be massive spheroid-dominated galaxies. Ultra-deep X-ray observations of SMGs undergoing major mergers have shown that AGN activity can be ongoing in both galactic components¹². Presumably in these major-merger events the black holes will eventually coalesce, further increasing the mass of the black hole in the resultant galaxy. □

Received 12 October 2004; accepted 11 February 2005; doi:10.1038/nature03473.

1. Tremaine, S. *et al.* The slope of the black hole mass versus velocity dispersion correlation. *Astrophys. J.* **574**, 740–753 (2002).
2. Yu, Q. & Tremaine, S. Observational constraints on the growth of massive black holes. *Mon. Not. R. Astron. Soc.* **335**, 965–976 (2002).
3. Barger, A. J. *et al.* The cosmic evolution of hard X-ray selected active galactic nuclei. *Astron. J.* **129**, 578–609 (2005).
4. McLure, R. J. & Dunlop, J. S. The cosmological evolution of quasar black hole masses. *Mon. Not. R. Astron. Soc.* **352**, 1390–1404 (2004).
5. Page, M. J., Stevens, J. A., Ivison, R. J. & Carrera, F. J. The evolutionary sequence of active galactic nuclei and galaxy formation revealed. *Astrophys. J.* **611**, L85–L88 (2004).
6. Smail, I., Ivison, R. J., Blain, A. W. & Kneib, J.-P. The nature of faint submillimetre-selected galaxies. *Mon. Not. R. Astron. Soc.* **331**, 495–520 (2002).
7. Chapman, S. C., Blain, A. W., Ivison, R. J. & Smail, I. R. A median redshift of 2.4 for galaxies bright at submillimetre wavelengths. *Nature* **422**, 695–698 (2003).
8. Hughes, D. H. *et al.* High-redshift star formation in the Hubble Deep Field revealed by a submillimetre-wavelength survey. *Nature* **394**, 241–247 (1998).
9. Chapman, S. C., Blain, A. W., Smail, I. R. & Ivison, R. J. A redshift survey of the submillimetre galaxy population. *Astrophys. J.* **622**, 2–26 (2005).
10. Stevens, J. A., Page, M. J., Ivison, R. J., Smail, I. & Carrera, F. J. A filamentary structure of massive star-forming galaxies associated with an X-ray-absorbed QSO at $z = 1.8$. *Astrophys. J.* **604**, L17–L20 (2004).
11. Croom, S. M. *et al.* The 2dF QSO Redshift Survey – XII. The spectroscopic catalogue and luminosity function. *Mon. Not. R. Astron. Soc.* **349**, 1397–1418 (2004).
12. Alexander, D. M. *et al.* The Chandra Deep Field North Survey. XIV. X-ray-detected obscured AGNs and starburst galaxies in the bright submillimetre source population. *Astron. J.* **125**, 383–397 (2003).
13. Ivison, R. J. *et al.* Spitzer observations of MAMBO galaxies: weeding out active nuclei in starbursting proto-ellipticals. *Astrophys. J. Suppl.* **154**, 124–129 (2004).
14. Greve, T. R. *et al.* An interferometric CO survey of luminous submm galaxies. *Mon. Not. R. Astron. Soc.* (in the press); preprint at (<http://arxiv.org/astro-ph/0503055>) (2005).
15. Tecza, M. *et al.* SPIFI observations of the starburst SMM J14011+0252: Already old, fat, and rich by $z = 2.565$. *Astrophys. J.* **605**, L109–L112 (2004).
16. Swinbank, A. M. *et al.* The rest-frame optical spectra of SCUBA galaxies. *Astrophys. J.* **617**, 64–80 (2004).
17. Smail, I., Chapman, S. C., Blain, A. W. & Ivison, R. J. The rest-frame optical properties of SCUBA galaxies. *Astrophys. J.* **616**, 71–85 (2004).
18. Alexander, D. M. *et al.* The Chandra Deep Field North Survey. XIII. 2 Ms point-source catalogs. *Astron. J.* **126**, 539–574 (2003).
19. Alexander, D. M. *et al.* The X-ray properties of SCUBA galaxies. *Astrophys. J.* (submitted).
20. Veilleux, S., Kim, D.-C. & Sanders, D. B. Optical spectroscopy of the IRAS 1 Jy sample of ultraluminous infrared galaxies. *Astrophys. J.* **522**, 113–138 (1999).
21. Steidel, C. C. *et al.* A survey of star-forming galaxies in the $1.4 < z < 2.5$ redshift desert: Overview. *Astrophys. J.* **604**, 534–550 (2004).
22. Fadda, D. *et al.* The AGN contribution to mid-infrared surveys. X-ray counterparts of the mid-IR sources in the Lockman Hole and HDF-N. *Astron. Astrophys.* **383**, 838–853 (2002).
23. Fabian, A. C. The obscured growth of massive black holes. *Mon. Not. R. Astron. Soc.* **308**, L39–L43 (1999).
24. Archibald, E. N. *et al.* Coupled spheroid and black hole formation, and the multifrequency detectability of active galactic nuclei and submillimetre sources. *Mon. Not. R. Astron. Soc.* **336**, 353–362 (2002).
25. Granato, G. L., De Zotti, G., Silva, L., Bressan, A. & Danese, L. A physical model for the coevolution of QSOs and their spheroidal hosts. *Astrophys. J.* **600**, 580–594 (2004).
26. Dunlop, J. S. *et al.* Quasars, their host galaxies and their central black holes. *Mon. Not. R. Astron. Soc.* **340**, 1095–1135 (2003).
27. Conselice, C. J., Chapman, S. C. & Windhorst, R. A. Evidence for a major merger origin of high-redshift submillimetre galaxies. *Astrophys. J.* **596**, L5–L8 (2003).
28. Springel, V., Di Matteo, T. & Hernquist, L. Black holes in galaxy mergers: the formation of red elliptical galaxies. *Astrophys. J.* **620**, L79–L82 (2005).
29. Elvis, M. *et al.* Atlas of quasar energy distributions. *Astrophys. J. Suppl.* **95**, 1–68 (1994).
30. Marconi, A. *et al.* Local supermassive black holes, relics of active galactic nuclei and the X-ray background. *Mon. Not. R. Astron. Soc.* **351**, 169–185 (2004).

Acknowledgements We are grateful to R. McLure, M. Page, F. Shankar and Q. Yu for providing data and scientific insight. We thank the Royal Society (D.M.A., I.S.), PPARC (F.E.B.) and NASA (S.C.C., W.N.B.) for support. Data presented here were obtained using the W.M. Keck Observatory, which is operated as a scientific partnership among Caltech, the University of California and NASA.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.M.A. (dma@ast.cam.ac.uk).

Infrared radiation from an extrasolar planet

Drake Deming¹, Sara Seager³, L. Jeremy Richardson² & Joseph Harrington⁴

¹Planetary Systems Laboratory and Goddard Center for Astrobiology, Code 693;

²Exoplanet and Stellar Astrophysics Laboratory, Code 667, NASA's Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

³Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road NW, Washington DC 20015, USA

⁴Center for Radiophysics and Space Research, Cornell University, 326 Space Sciences Bldg, Ithaca, New York 14853-6801, USA

A class of extrasolar giant planets—the so-called ‘hot Jupiters’ (ref. 1)—orbit within 0.05 AU of their primary stars (1 AU is the Sun–Earth distance). These planets should be hot and so emit detectable infrared radiation². The planet HD 209458b (refs 3, 4) is an ideal candidate for the detection and characterization of this infrared light because it is eclipsed by the star. This planet has an anomalously large radius (1.35 times that of Jupiter³), which may be the result of ongoing tidal dissipation⁶, but this explanation requires a non-zero orbital eccentricity (~ 0.03 ; refs 6, 7), maintained by interaction with a hypothetical second planet. Here we report detection of infrared (24 μ m) radiation from HD 209458b, by observing the decrement in flux during secondary eclipse, when the planet passes behind the star. The planet's 24- μ m flux is 55 ± 10 mJy (1 σ), with a brightness temperature of $1,130 \pm 150$ K, confirming the predicted heating by stellar irradiation^{2,8}. The secondary eclipse occurs at the midpoint between transits of the planet in front of the star (to within ± 7 min, 1 σ), which means that a dynamically significant orbital eccentricity is unlikely.

Operating cryogenically in a thermally stable space environment, the Spitzer Space Telescope⁹ has sufficient sensitivity to detect hot Jupiters at their predicted infrared flux levels⁸. We observed the secondary eclipse (hereafter referred to as ‘the eclipse’) of HD 209458b with the 24- μ m channel of the Multiband Imaging Photometer for Spitzer (MIPS)¹⁰. Our photometric time series observations began on 6 December 2004 at 21:29 UTC (Coordinated Universal Time), and ended at approximately 03:23 UTC on 7 December 2004 (5 h 54 min duration). We analyse 1,696 of the 1,728 10-s exposures so acquired, rejecting 32 images having obvious flaws. The Supplementary Information contains a sample image, together with information on the noise properties of the data.

We first verify that circumstellar dust does not contribute significantly to the stellar flux. Summing each stellar image over a 13×13 pixel synthetic aperture (33×33 arcsec), we multiply the average sum by 1.15 to account for the far wings of the point spread function (PSF)¹¹, deriving a flux of 21.17 ± 0.11 mJy. The temperature of the star is close to 6,000 K (ref. 12). At a distance of 47 pc (ref. 13), a model atmosphere¹⁴ predicts a flux of 22 mJy, agreeing

with our observed flux to within an estimated ~ 2 mJy error in absolute calibration. We conclude that the observed flux is dominated by photospheric emission, in agreement with a large Spitzer study of planet-bearing stars at this wavelength¹¹.

Our time series analysis is optimized for high relative precision. We extract the intensity of the star from each image using optimal photometry with a spatial weighting function¹⁵. Selecting the Tiny Tim¹⁶ synthetic MIPS PSF for a 5,000-K source at $24\text{ }\mu\text{m}$, we spline-interpolate it to 0.01 pixel spacing, rebin it to the data resolution, and centre it on the stellar image. The best centring is judged by a least-squares fit to the star, fitting to within the noise level. The best-centred PSF becomes the weighting function in deriving the stellar photometric intensity. We subtract the average background over each image before applying the weights. MIPS data includes per-pixel error estimates¹⁷, which we use in the optimal photometry and to compute errors for each photometric point. The optimal algorithm¹⁵ predicts the signal-to-noise ratio (SNR) for each photometric point, and these average to 119. Our data are divided into 14 blocks, defined by pre-determined raster positions of the star on the detector. To check our SNR, we compute the internal

scatter within each block. This gives SNR in the range from 95 to 120 (averaging 111), in excellent agreement with the optimal algorithm. For each point we use the most conservative possible error: either the scatter within that block or the algorithm estimate, whichever is greater. We search for correlations between the photometric intensities and small fluctuations in stellar position, but find none. We also perform simple aperture photometry on the images, and this independent procedure confirms our results, but with 60% greater errors.

The performance of MIPS at $24\text{ }\mu\text{m}$ is known to be excellent¹⁸. Only one instrument quirk affects our photometry. The MIPS observing sequence obtains periodic bias images, which reset the detector. Images following resets have lower overall intensities (by $\sim 0.1\text{--}1\%$), which recover in later images. The change is common to all pixels in the detector, and we remove it by dividing the stellar intensities by the average zodiacal background in each image. We thereby remove variations in instrument/detector response, both known and unknown. The best available zodiacal model¹⁹ predicts a background increase of 0.18% during the ~ 6 h of our photometry. Because the star will not share this increase, we remove a 0.18% linear baseline from the stellar photometry. Note that the eclipse involves both a decrease and increase in flux, and its detection is insensitive to monotonic linear baseline effects.

To detect weak signals reliably requires investigating the nature of the errors. We find that shot noise in the zodiacal background is the dominant source of error; systematic effects are undetectable after normalizing any individual pixel to the total zodiacal background. All of our results are based on analysis of the 1,696 individual photometric measurements versus heliocentric phase from a recent

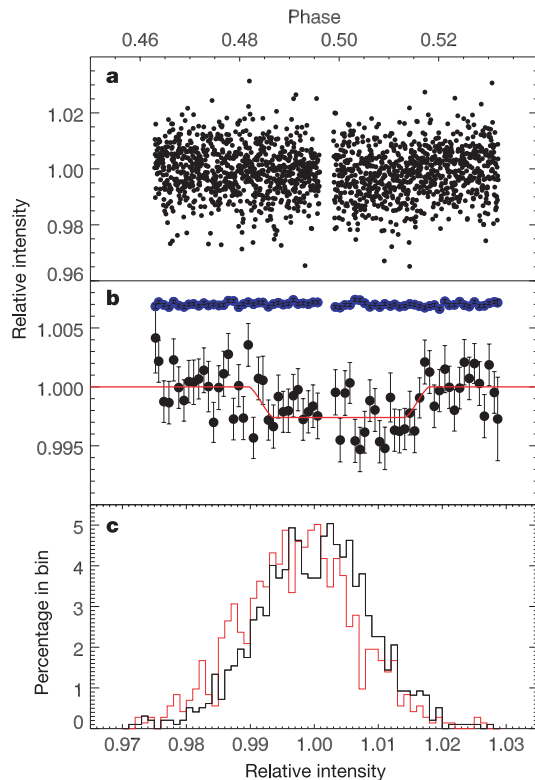


Figure 1 Observations showing our detection of the secondary eclipse in HD 209458. **a**, Relative intensities versus heliocentric phase (scale at top) for all 1,696 data points. The phase is corrected for light travel time at the orbital position of the telescope. Error bars are suppressed for clarity. The gap in the data near phase 0.497 is due to a pause for telescope overhead activity. The secondary eclipse is present, but is a factor of ~ 4 below the $\sim 1\%$ noise level of a single measurement. **b**, Intensities from **a**, averaged in bins of phase width 0.001 (scale at top), with 1σ error bars computed by statistical combination from the errors of individual points. The red line is the best-fit secondary eclipse curve (depth = 0.26%), constrained to a central phase of 0.5. The points in blue are a control sequence, summing intensities over a 10×10 -pixel region of the detector, to beat down the random errors and reveal any possible systematic effects. The control sequence uses the same detector pixels, on average, as those where the star resides, but is sampled out of phase with the variations in the star's raster motion during the MIPS photometry cycle. **c**, Histograms of intensity (lower abscissa scale) for the points in **a**, with bin width 0.1%, shown separately for the out-of-eclipse (black) and in-eclipse intervals (red).

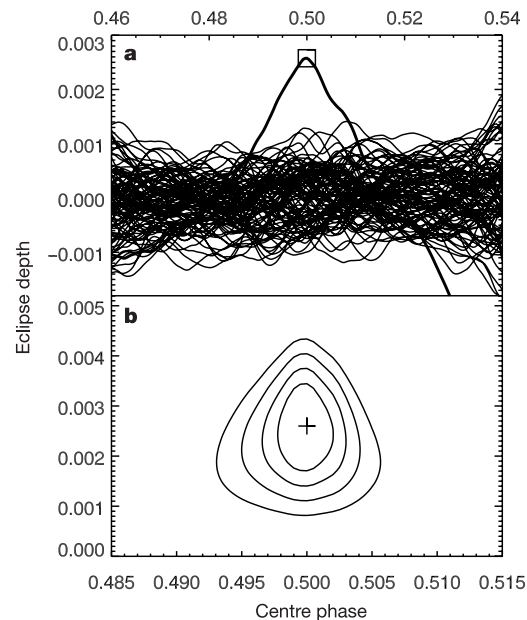


Figure 2 Amplitude of the secondary eclipse versus assumed central phase, with confidence intervals for both. **a**, The darkest line shows the amplitude of the best-fit eclipse curve versus the assumed central phase (scale at top). The overplotted point marks the fit having smallest χ^2 , which also has the greatest eclipse amplitude. The numerous thinner black lines show the effect of fitting to 100 synthetic data sets containing no eclipse, but having the same per-point errors as the real data. Their fluctuations in retrieved amplitude versus phase are indicative of the error in eclipse amplitude, and are consistent with $\sigma = 0.046\%$. Note that the eclipse amplitude found in the real data (0.26%) stands well above the error envelope at phase 0.5. **b**, Confidence intervals at the 1, 2, 3 and 4σ levels for the eclipse amplitude and central phase (note expansion of phase scale, at bottom). The plotted point marks the best fit (minimum χ^2) with eclipse depth of 0.26%, and central phase indistinguishable from 0.5. The centre of the eclipse occurs in our data at Julian day 2453346.5278.

ephemeris²⁰ (Fig. 1a). We propagate the individual errors (not shown on Fig. 1a) through a transit curve fit to calculate the error on the eclipse depth. Because about half of the 1,696 points are out of eclipse, and half are in eclipse, and the SNR ≈ 111 per point, the error on the eclipse depth should be $\sim 0.009 \times 2^{0.5}/848^{0.5} = 0.044\%$ of the stellar continuum. Model atmospheres for hot Jupiters^{2,8,21–24} predict eclipse depths in the range from 0.2–0.4% of the stellar continuum, so we anticipate a detection of 4–9 σ significance. The eclipse is difficult to discern by eye on Fig. 1a, because the observed depth (0.26%) is a factor of 4 below the scatter of individual points. We use the known period (3.524 days) and radii⁵ to fit an eclipse curve to the Fig. 1a data, varying only the eclipse depth, and constraining the central phase to 0.5. This fit detects the eclipse at a depth of $0.26\% \pm 0.046\%$, with a reduced χ^2 of 0.963, denoting a good fit. Note that the 5.6 σ significance applies to the aggregate result, not to individual points. The eclipse is more readily seen by eye on Fig. 1b, which presents binned data and the best-fit eclipse curve. The data are divided into many bins, so the aggregate 5.6 σ significance is much less for a single bin (SNR ≈ 1 per point). Nevertheless, the dip in flux due to the eclipse is apparent, and the observed duration is approximately as expected. As a check, we use a control photometric sequence (Fig. 1b) to eliminate false positive detection of the eclipse due to instrument effects. We also plot the distribution of points in intensity for both the in-eclipse and out-of-eclipse phase intervals (Fig. 1c). This shows that the entire distribution shifts as expected with the eclipse, providing additional discrimination against a false positive detection.

We further illustrate the reality of the eclipse on Fig. 2. Now shifting the eclipse curve in phase, we find the best-fitting amplitude and χ^2 at each shift. This determines the best-fit central phase for the eclipse, and also further illustrates the statistical significance of the result. The thick line in Fig. 2a shows that the maximum amplitude (0.26%) is obtained at exactly phase 0.5 (which is also the minimum of χ^2). Further, we plot the eclipse ‘amplitude’ versus central phase using 100 sets of synthetic data, consisting of gaussian noise with dispersion matching the real data, but without an eclipse. The amplitude (0.26%) of the eclipse in the real data stands well above

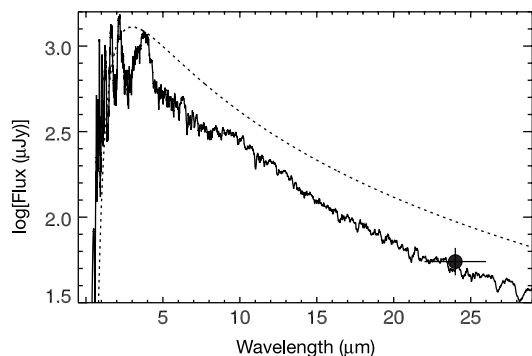


Figure 3 Flux from a model atmosphere shown in comparison to our measured infrared flux at 24 μm . A theoretical spectrum (solid line) shows that planetary emission (dominated by absorbed and re-radiated stellar radiation) should be very different from a blackbody. Hence, models are required to interpret the 24- μm flux measurement in terms of the planetary temperature. The model shown has $T_{\text{eq}} = 1,700\text{ K}$ and was computed from a one-dimensional plane-parallel radiative transfer model, considering a solar system abundance of gases, no clouds, and the absorbed stellar radiation re-emitted on the day side only. Note the marked difference from a 1,700-K blackbody (dashed line), although the total flux integrated over the blackbody spectrum is equal to the total flux integrated over the model spectrum. (The peaks at short wavelength dominate the flux integral in the atmosphere model, note log scale in the ordinate.) The suppressed flux at 24 μm is due to water vapour opacity. This model lies at the hot end of the range of plausible models consistent with our measurement, but the error bars admit models with cooler T_{eq} .

the statistical fluctuations in the synthetic data.

Figure 2b shows confidence intervals on the amplitude and central phase, based on the χ^2 values. The phase shift of the eclipse is quite sensitive to eccentricity (e) and is given²⁵ as $\Delta t = 2Pe\cos(\omega)/\pi$, where P is the orbital period, and ω is the longitude of periastron. The Doppler data alone give $e = 0.027 \pm 0.015$ (Laughlin, G., personal communication), and allow a phase shift as large as ± 0.017 (87 min). We find the eclipse centred at phase 0.5, and we checked the precision using a bootstrap Monte Carlo procedure²⁶. The 1 σ phase error from this method is 0.0015 (~ 7 min), consistent with Fig. 2b. A dynamically significant eccentricity, $e \approx 0.03$ (refs 6, 7), constrained by our 3 σ limit of $\Delta t < 21$ min, requires $|(\omega - \pi/2)| < 12$ degrees and is therefore only possible in the unlikely case that our viewing angle is closely parallel to the major axis of the orbit. A circular orbit rules out a promising explanation for the planet's anomalously large radius: tidal dissipation as an interior energy source to slow down planetary evolution and contraction⁷. Because the dynamical time for tidal decay to a circular orbit is short, this scenario posited the presence of a perturbing second planet in the system to continually force the eccentricity—a planet that is no longer necessary with a circular orbit for HD209458b.

The infrared flux from the planet follows directly from our measured stellar flux (21.2 mJy) and the eclipse depth (0.26%), giving $55 \pm 10\ \mu\text{Jy}$. The error is dominated by uncertainty in the eclipse depth. Using the planet's known radius⁵ and distance¹³, we obtain a brightness temperature $T_{24} = 1,130 \pm 150\text{ K}$, confirming heating by stellar irradiation². Nevertheless, T_{24} could differ significantly from the temperature of the equivalent blackbody (T_{eq}), that is, one whose bolometric flux is the same as the planet. Without measurements at shorter wavelengths, a model atmosphere must be used to estimate T_{eq} from the 24- μm flux. One such model is shown in Fig. 3, having $T_{\text{eq}} = 1,700\text{ K}$. This temperature is much higher than T_{24} (1,130 K) due to strong, continuous H_2O vapour absorption at 24 μm . The bulk of the planetary thermal emission derives ultimately from re-radiated stellar irradiation, and is emitted at 1–4 μm , between H_2O bands. However, our 24- μm flux error admits a range of models, including some with a significantly lower T_{eq} (for example, but not limited to, models with reflective clouds or less H_2O vapour).

Shortly after submission of this Letter, we became aware of a similar detection for the TrES-1 transiting planet system²⁷ using Spitzer's Infrared Array Camera²⁸. Together, these Spitzer results represent the first measurement of radiation from extrasolar planets. Additional Spitzer observations should rapidly narrow the range of acceptable models, and reveal the atmospheric structure, composition, and other characteristics of close-in extrasolar giant planets. □

Received 3 February; accepted 28 February 2005; doi:10.1038/nature03507.

Published online 23 March 2005.

- Collier-Cameron, A. Extrasolar planets: what are hot Jupiters made of? *Astron. Geophys.* **43**, 421–425 (2002).
- Seager, S. & Sasselov, D. D. Extrasolar giant planets under strong stellar irradiation. *Astrophys. J.* **502**, L157–L161 (1998).
- Charbonneau, D., Brown, T. M., Latham, D. W. & Mayor, M. Detection of planetary transits across a sun-like star. *Astrophys. J.* **529**, L45–L48 (2000).
- Henry, G. W., Marcy, G. W., Butler, R. P. & Vogt, S. S. A transiting “51 Peg-like” planet. *Astrophys. J.* **529**, L41–L44 (2000).
- Brown, T. M., Charbonneau, D., Gilliland, R. L., Noyes, R. W. & Burrows, A. Hubble Space Telescope time-series photometry of the transiting planet of HD 209458. *Astrophys. J.* **552**, 699–709 (2001).
- Bodenheimer, P., Lin, D. N. C. & Mardling, R. A. On the tidal inflation of short-period extrasolar planets. *Astrophys. J.* **548**, 466–472 (2001).
- Laughlin, G. et al. A comparison of observationally determined radii with theoretical radius predictions for short-period transiting extrasolar planets. *Astrophys. J.* (in the press).
- Burrows, A., Sudarsky, D. & Hubeny, I. In *The Search for Other Worlds: Proc. 14th Annu. Astrophys. Conf. in Maryland* (eds Holt, S. & Deming, D.) Vol. 713, 143–150 (American Institute of Physics, Melville, New York, 2003).
- Werner, M. W. et al. The Spitzer Space Telescope mission. *Astrophys. J. Suppl.* **154**, 1–9 (2004).
- Rieke, G. H. et al. The Multiband Imaging Photometer for Spitzer (MIPS). *Astrophys. J. Suppl.* **154**, 25–29 (2004).

11. Beichman, C. A. *et al.* Planets and IR excesses: preliminary results from a *Spitzer* MIPS survey of solar-type stars. *Astrophys. J.* (in the press).
12. Ribas, A. H., Solano, E., Masana, E. & Gimenez, A. Effective temperatures and radii of planet-hosting stars from IR photometry. *Astron. Astrophys.* **411**, L501–L504 (2003).
13. Perryman, M. A. C. (ed.) *The Hipparcos and Tycho Catalogues* (ESA SP-1200, European Space Agency, Noordwijk, 1997).
14. Kurucz, R. *Solar Abundance Model Atmospheres for 0, 1, 2, 4, and 8 km/s* CD-ROM 19 (Smithsonian Astrophysical Observatory, Cambridge, Massachusetts, 1994).
15. Horne, K. An optimal extraction algorithm for CCD spectroscopy. *Publ. Astron. Soc. Pacif.* **98**, 609–617 (1986).
16. Krist, J. in *Astronomical Data Analysis Software and Systems IV* (eds Shaw, R. A., Payne, H. E. & Hayes, J. J. E.) Vol. 77, 349–352 (Astronomical Society of the Pacific, San Francisco, 1995).
17. Gordon, K. D. Reduction algorithms for the Multiband Imaging Photometer for SIRTF. *Publ. Astron. Soc. Pacif.* (in the press).
18. Rieke, G. H. *et al.* in *Proc. SPIE: Optical, Infrared, and Millimeter Space Telescopes* (ed. Mather, J. C.) Vol. 5487, 50–61 (SPIE, Bellingham, Washington, 2004).
19. Kelsall, T. *et al.* The COBE Diffuse Infrared Background Experiment (DIRBE) search for the cosmic infrared background. II. Model of the interplanetary dust cloud. *Astrophys. J.* **508**, 44–73 (1998).
20. Wittenmyer, R. A. *The Orbital Ephemeris of HD 209458b*. Master's thesis, San Diego State Univ. (2003).
21. Goukenleque, C., Bezaud, B., Joguet, B., Lellouch, E. & Freedman, R. A radiative equilibrium model of 51 Peg b. *Icarus* **143**, 308–323 (2000).
22. Seager, S., Whitney, B. A. & Sasselov, D. D. Photometric light curves and polarization of close-in extrasolar giant planets. *Astrophys. J.* **540**, 504–520 (2000).
23. Barman, T. S., Hauschildt, P. H. & Allard, F. Irradiated planets. *Astrophys. J.* **556**, 885–895 (2001).
24. Sudarsky, D., Burrows, A. & Hubeny, I. Theoretical spectra and atmospheres of extrasolar giant planets. *Astrophys. J.* **588**, 1121–1148 (2003).
25. Charbonneau, D. in *Scientific Frontiers in Research on Extrasolar Planets* (eds Deming, D. & Seager, S.) Vol. 294, 449–456 (Astronomical Society of the Pacific, San Francisco, 2003).
26. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical recipes in C* 2nd edn (Cambridge Univ. Press, Cambridge, 1992).
27. Alonso, R. *et al.* TrES-1, the transiting planet of a bright K0V star. *Astrophys. J.* **613**, L153–L156 (2004).
28. Charbonneau, D. *et al.* Detection of thermal emission from an extrasolar planet. *Astrophys. J.* (in the press).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank G. Laughlin for communicating the latest orbital eccentricity solutions from the Doppler data and for his evaluation of their status. We acknowledge informative conversations with D. Charbonneau, G. Marcy, B. Hansen, K. Menou and J. Cho. This work is based on observations made with the *Spitzer* Space Telescope, which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract to NASA. Support for this work was provided directly by NASA, and by its Origins of Solar Systems programme and Astrobiology Institute. We thank all the personnel of the *Spitzer* telescope and the MIPS instrument, who ultimately made these measurements possible. L.J.R. is a National Research Council Associate at NASA's Goddard Space Flight Center.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.D. (Leo.D.Deming@nasa.gov).

Tuning clathrate hydrates for hydrogen storage

Huen Lee¹, Jong-won Lee¹, Do Youn Kim¹, Jeasung Park¹, Yu-Taek Seo^{2*}, Huang Zeng², Igor L. Moudrakovski², Christopher I. Ratcliffe² & John A. Ripmeester²

¹Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Republic of Korea
²Steele Institute for Molecular Sciences, National Research Council Canada, Ottawa, Ontario, Canada K1A 0R6

* Present address: Conversion Process Research Center, Korea Institute of Energy Research, PO Box 103, Jang-dong, Yuseong-gu, Daejeon 305-343, Republic of Korea

The storage of large quantities of hydrogen at safe pressures¹ is a key factor in establishing a hydrogen-based economy. Previous strategies—where hydrogen has been bound chemically², adsorbed in materials with permanent void space³ or stored in hybrid materials that combine these elements³—have problems arising from either technical considerations or materials cost^{2–5}. A recently reported^{6–8} clathrate hydrate of hydrogen exhibiting

two different-sized cages does seem to meet the necessary storage requirements; however, the extreme pressures (~2 kbar) required to produce the material make it impractical. The synthesis pressure can be decreased by filling the larger cavity with tetrahydrofuran (THF) to stabilize the material⁹, but the potential storage capacity of the material is compromised with this approach. Here we report that hydrogen storage capacities in THF-containing binary-clathrate hydrates can be increased to ~4 wt% at modest pressures by tuning their composition to allow the hydrogen guests to enter both the larger and the smaller cages, while retaining low-pressure stability. The tuning mechanism is quite general and convenient, using water-soluble hydrate promoters and various small gaseous guests.

The structure II (sII) hydrates constitute a large family of clathrates with an ideal unit cell 16S·8L·136H₂O, where the large (L) and small (S) cavities can be filled with guest molecules¹⁰. The 'solid solution' theory of van der Waals and Platteeuw, the classical approach to understanding clathrate behaviour, uses the basic expression:

$$\Delta\mu_w = -kT\Sigma_i\nu_i \ln(1 - \Theta_i) \quad (1)$$

which relates the free energy difference, $\Delta\mu_w$, between ice and a hypothetical empty hydrate framework to the minimum occupancy Θ_i of the hydrate cavities of type i required to render it stable; ν_i are the number of cages of type i in the unit cell normalized by the number of water molecules. For sII hydrate, the 'best value' of $\Delta\mu_w$ of 884 J mol⁻¹ requires the large cages to be filled to more than 99% for stability. This is consistent with the observation that many sII hydrates are known for which the stoichiometry is L·17H₂O within experimental error. The small cages can be occupied by a small guest, leading to a double hydrate, (2S)_x·L·17H₂O, generally stable to higher temperatures, where x is the fractional occupancy of the small cages, as recently illustrated⁹ for THF and H₂ ($x \approx 1$, 1 wt% H₂). The recently reported H₂ clathrate is also sII with multiple occupancy of the cages (4 in L, 2 in S)^{6–8}. Double hydrates represent an opportunity to engineer hybrid structures that combine H₂ storage capacity with much less severe synthetic pressures.

We carried out initial studies on materials produced from 5.56 mol% THF solution in water, which gives a clathrate of composition THF·17H₂O when cooled below the melting point of 277.3 K. The THF hydrate was then pressurized with H₂ gas at various pressures up to ~120 bar, and examined for structure, cage populations and composition by powder X-ray diffraction (PXRD), Raman and NMR spectroscopy, and direct measurement of the H₂ released on decomposition. From the PXRD results (Supplementary Fig. 1 and Supplementary Table 1), the material was confirmed to be a sII hydrate according to its phase behaviour (Supplementary Fig. 2). The Raman spectra (Fig. 1) show four transitions due to rotational fine structure associated with the high-pressure H₂ gas, and a broad band that can be identified with H₂ inside hydrate cages. The hydrate H₂ line is shifted to lower frequency compared to the free gas¹¹, as has also been observed for O₂ and N₂ hydrates¹². ¹H NMR spectroscopy was used to monitor H₂ in the product of the reaction of H₂ with perdeuterated THF hydrate (THF-d₈·17D₂O), so that only H₂ signals and residual protons in water and THF would be observed. The spectra in Fig. 2 show a broad line at 4.3 p.p.m. that can be attributed to H₂ in the small cavities of the double hydrate. Full loading of the small cavities of THF·17H₂O with hydrogen (2H₂ per small cage) will result in a storage capacity of 2.1 wt% H₂.

Is it possible to obtain higher hydrogen loading, while keeping the H₂ pressure at a reasonable value? Although double hydrates have been known for over a century, techniques for hydrate analysis in terms of guest distribution over the hydrate cage sites were developed only recently, so little effort has gone into attempts to tune hydrate compositions. In order to increase the hydrogen content of the hydrate, the hydrogen guest must also enter the

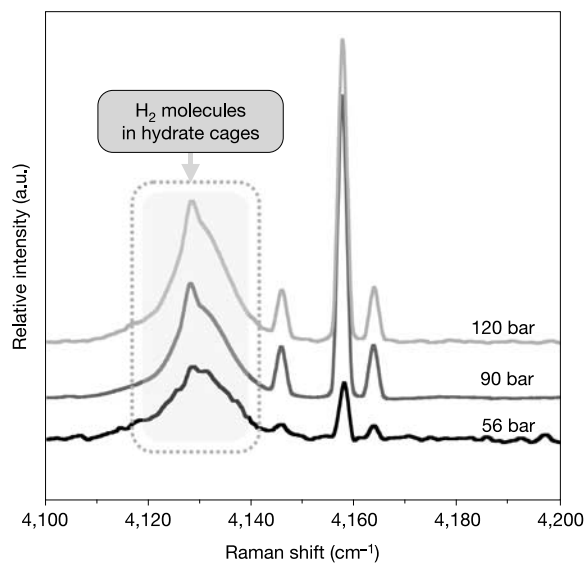


Figure 1 Raman spectra of the THF + H₂ double hydrates. Frozen 5.56 mol% THF solutions were stored in a refrigerator at 243 K for at least one day, and then ground to a fine powder (~200 μm). The powdered solid was exposed to hydrogen gas (ultrapure carrier grade, 99.999%) at pressures from 56 to 120 bar and at a temperature of 270 K in a high-pressure Raman cell equipped with two sapphire windows and circular grooves for coolant. The Raman spectrum was obtained using a SPEX 1404p single grating Raman spectrometer, with a CCD detector cooled by liquid nitrogen, using the focused 488 nm line of an Ar-ion laser for excitation. The laser intensity was typically 300 mW. The H–H vibron peak at 4,128 cm^{−1} consists of both the sharp spectral line assigned to hydrogen gas and the broad peak due to hydrogen molecules stored in clathrate hydrate cages. The broad peak becomes higher when hydrate formation pressure increases in the stable region of the pressure–temperature diagram (Supplementary Fig. 2).

large cavities of sII. The solid-solution model then requires that the large cage occupancies $\theta_L(\text{THF})$ and $\theta_L(\text{H}_2)$ become comparable, where the θ_L are given by $\theta_{Li} = C_{Li}p_i / (1 + C_{L1}p_1 + C_{L2}p_2)$; the C_{Li} are the Langmuir constants for the large cage, and the p_i are the partial pressures of the guest. As $C_L(\text{THF})$ is much larger than $C_L(\text{H}_2)$, this requires the partial pressure (concentration) of THF to be lowered considerably in order to place H₂ clusters into the large cage.

Hence, a number of experiments were carried out at an H₂ pressure of 120 bar where the THF concentration was decreased from 5.56 mol% down to 0.1 mol%. The compositions of the hydrates produced were obtained from measurements of the Raman band intensities. The H₂ content of the clathrate produced at each THF loading was calculated from the integral intensity of its Raman band (after appropriate subtraction of intensity from the overlapping gas lines) relative to that of the 5.56 mol% sample and

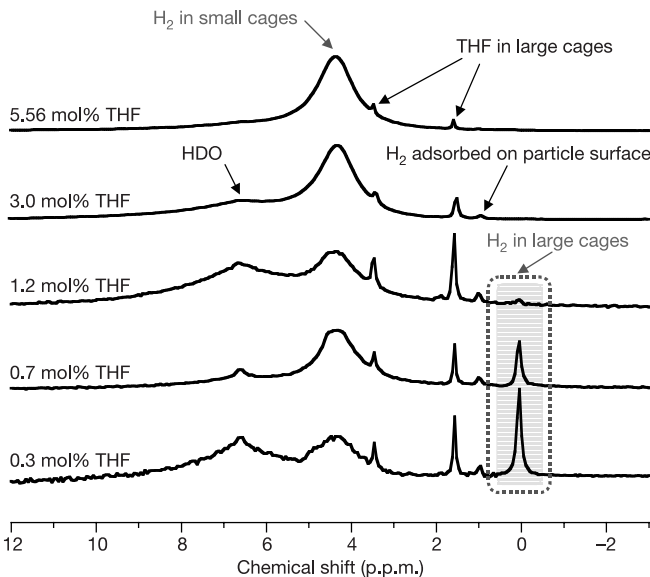


Figure 2 Magic angle spinning ¹H NMR spectra of the THF + H₂ double hydrates formed at 120 bar and 270 K as a function of concentration of THF. The NMR samples were prepared from deuterated water (D₂O, 99.9 at.% D) and THF (THF-d₈, 99.5 at.% D) by the same procedure as for Raman experiments. The chemical shifts of D₂O and THF-d₈ were identified from the NMR spectrum of THF-d₈ hydrate. Spectra were recorded using a Varian INOVA600 spectrometer, spin rate ~12 kHz, pulse length 5 μs, repetition time 15 s. The samples were transferred to an NMR rotor and analysed at 1 bar and 183 K. As the hydrate is not absolutely stable under these conditions, quantitative spectra cannot be obtained because of some decomposition, although a consistent picture can be constructed from the spectra that illustrate several features. The distribution of hydrogen molecules in both cages depends on temperature and pressure¹³.

taking into account the different relative volumes of hydrate and ice. Extinction coefficients were assumed to be constant, based on the observation that the ratio of Raman intensity to volumetrically measured H₂ content remains constant over the range of THF concentrations (Supplementary Material and Supplementary Fig. 3). The 5.56 mol% sample was taken to be fully loaded, with 2H₂ per small cage and 2.09 wt% H₂ content. Results are presented in Table 1 as H₂:THF mole fractions and wt% H₂ content of the hydrate phase (based on the model discussed below). The H₂:THF mole ratio of the ‘reference’ 5.56 mol% sample is 4, whereas for samples with 4.0 and 2.0 mol% THF the mole ratio falls slightly below 4, indicating a small deficit of H₂ in the small cages.

What is most interesting, however, is that at 1.0 mol% and down to 0.15 mol% the mole ratio increases above 4 to as high as ~23 for the 0.15 mol% sample, indicating that large cages must also contain H₂. The 1 mol% concentration of THF is approximately at the

Table 1 Raman peak areas and H₂ content parameters

Mol% THF (starting solution)	Raman peak area (a.u.)	H ₂ :THF mole ratio in hydrate	x	y	Max. wt% H ₂ in hydrate
10.0	15,796	4.0	0	2	2.09*
5.56	15,406	4.0	0	2	2.09
4.0	9,740	3.44	0	1.719	1.80
2.0	4,896	3.36	0	1.681	1.76
1.5	4,428	4.03	0.003	2	2.09
1.0	3,330	4.52	0.061	2	2.24
0.5	3,085	8.35	0.352	2	3.00
0.2	2,613	17.63	0.630	2	3.80
0.15	2,568	23.10	0.705	2	4.03
0.1	0	0	0	0	0

For (yH₂)₂·(4H₂)_x·THF_(1-x)·17H₂O. a.u., arbitrary units.

*This sample has excess THF (with excess H₂ dissolved in the THF). Note that pure hydrogen clathrate (2H₂)₂·(4H₂)·17H₂O would have a 5.002 wt% H₂ content.

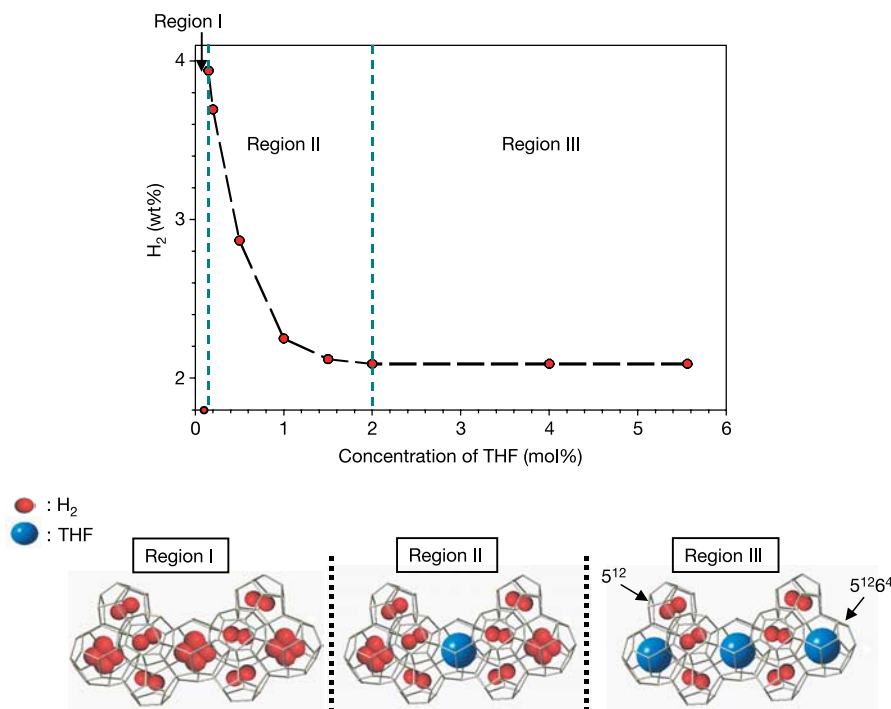
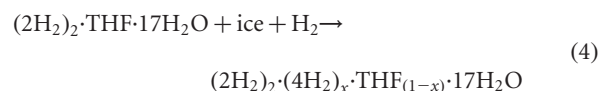


Figure 3 H₂ gas content as a function of THF concentration, and a schematic diagram of H₂ distribution in the cages of THF + H₂ hydrate. (H₂ gas content is calculated from g of H₂ per g of hydrate, and expressed as wt%). In region III, H₂ molecules are only stored in small cages, while in region II both small and large cages can store H₂ molecules. At the

highly dilute THF concentrations of region I, H₂ molecules can still be stored in both cages, but extreme pressures (~2 kbar) are required to form the hydrates. Pure H₂ clathrate (2H₂)₂·(4H₂)·17H₂O would have a 5.002 wt% H₂ content.

eutectic composition of THF hydrate and below this, solid ice I_h would be the first phase to appear, thus reflecting the fact that the concentration of THF has declined to the point where hydrate no longer forms ($\Theta_L = C_L p / (1 + C_L p) < 1$). For sII hydrate to be stable relative to ice, the only condition that needs to be satisfied is that the occupancy of the large cage is close to 1, so $\Theta_{\text{THF}} + \Theta_{\text{H}_2} \approx 1$. This approach represents a general strategy for tuning hydrate compositions—at or below the eutectic composition of water-soluble sII hydrate formers, it is necessary to introduce other guests into the large cages of sII hydrate for stability.

At this stage we can consider the exact nature of the THF hydrate with high H₂ content. We have concluded that some large cages contain THF and some have H₂ clusters, while maintaining the double cage occupancy of the small cages. Thus the formation of double hydrate at low THF concentration (1.0–0.15 mol%) probably progresses as follows:



where (2H₂)₂ and (4H₂)_x represent clusters of two or four H₂ molecules occupying small cages and large cages, respectively (as suggested from very-high-pressure studies of H₂ hydrate⁶). THF hydrate formed in the first process (equation (2)) reacts with H₂ to form a double hydrate (equation (3)), which is the kinetically limited product. Disproportionation of the double hydrate formed in equation (3) with the ice must then occur in order to allow H₂ to occupy large cages and thus produce a thermodynamically stable double hydrate (equation (4)). ¹H NMR spectroscopy was used to identify the distribution of guests over the cages in the double

hydrate with high H₂ content. When the concentration of THF reaches ~1 mol% a new line appears at ~0.15 p.p.m., which continues to grow with decreasing THF concentration. Considering the other experimental evidence for the increased loading of the sample, we assign this peak to H₂ clusters in the large cages (Fig. 2).

We can then calculate the values of *x* in equation (4) and the effective wt% H₂ for the clathrate formed from each THF concentration. Figure 3 shows a plot of wt% H₂ stored in hydrate as a function of THF concentration—note that these represent a maximum wt% H₂ content, based on the assumptions above. Cross-checks were made with a direct gas release measurement for the 5.56, 0.7 and 0.5 mol% samples, which gave respectively 2.1, 2.4 and 2.7 wt% H₂ compared to 2.1, 2.6 and 3.0 wt% H₂ from the Raman measurements. This can explain the increase in H₂ content in the hydrate, and indeed we observe that the wt% H₂ was 2.24 at 1.0 mol% increasing to 4.03 wt% at 0.15 mol% THF. At even lower THF concentration, 0.1 mol%, a H₂-containing hydrate was no longer formed under the conditions used. We take this to mean that the combination of THF concentration and H₂ pressure was insufficient to fill the large cages in the structure to produce a stable hydrate. When H₂ was applied to frozen 0.2 mol% THF solutions at pressures between 90 and 140 bar, the H₂ content reached a steady state for pressures higher than ~100 bar. If the H₂ in the large cages is present in the form of (H₂)₄ clusters, as has been proposed for pure H₂ hydrate, about 60% of the large cages will be occupied by these, with the remainder expected to contain THF. Similar experiments have shown that, upon decreasing the THF concentration below 1.0 mol%, it is possible to introduce other guests such as CH₄ into the large cages of sII hydrate. In recent ¹³C NMR studies of the sII double hydrate of CH₄ and THF, two CH₄ resonances at −4.5 p.p.m. (small cage) and −8.1 p.p.m. (large cage) were found, but the large cage resonance was only seen in solutions with concentrations below 1.0 mol% THF.

Although the hydrate lattice appears to be tunable for higher

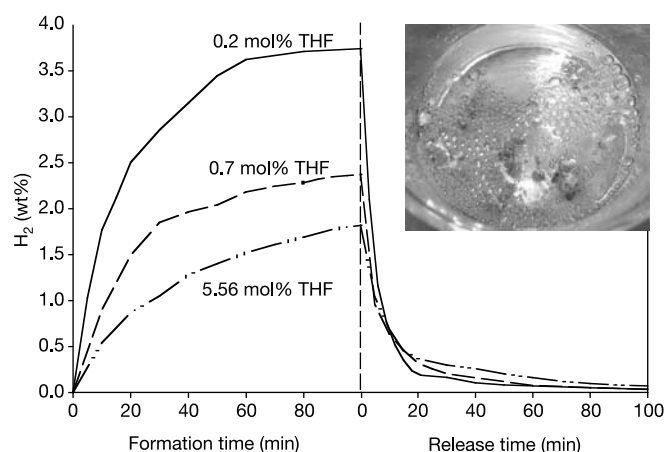


Figure 4 Formation/release kinetics of the H_2 + THF double hydrate in the pores of silica beads. At three THF concentrations (5.56, 0.7 and 0.2 mol%) the amount of H_2 stored in the cages was measured using an isometric micro-syringe pump. Hydrates were maintained at 1 bar and 270 K for complete release of hydrogen after forming at 120 bar and 270 K. Hydrogen bubbles are observed during hydrate release (inset).

levels of storage, up to ~ 4.0 wt%, the reaction is so slow as to make the synthetic approach impractical. It is beyond the scope of this Letter to devise an industrial strategy for the efficient synthesis of H_2 -containing hydrate, but a promising approach is as follows: The reaction of the hydrate/ice mixture with H_2 is limited by diffusion through a bulk solid phase. So, if the reacting phase can be dispersed so as to improve the surface area to volume ratio, improved kinetics should be possible. To demonstrate proof of principle, we have dispersed the aqueous phase on a silica bead support, and after exposure to H_2 gas, followed the reaction kinetics by measuring the rate at which gas is consumed. Figure 4 shows that the reaction proceeds as expected and that the conversion is now complete in ~ 1 h; this is a tremendous improvement from the bulk reaction, which took a week or more. We recognize that this is not a practical solution, as the product resides inside the silica gel, and the additional weight takes the material out of the required range as a hydrogen storage material. However, this does show that a dispersed phase will react in a reasonable time, giving a product that is a suitable storage material. □

Received 20 September 2004; accepted 25 January 2005; doi:10.1038/nature03457.

- Berry, G. D. & Aceves, S. M. Onboard storage alternatives for hydrogen vehicles. *Energy Fuels* **12**, 49–55 (1998).
- Schlapbach, L. & Züttel, A. Hydrogen-storage materials for mobile applications. *Nature* **414**, 353–358 (2001).
- Weitkamp, J., Fritz, M. & Ernst, S. Zeolites as media for hydrogen storage. *Int. J. Hydrogen Energy* **20**, 967–970 (1995).
- Schimmel, H. G. *et al.* Hydrogen adsorption in carbon nanostructures: Comparison of nanotubes, fibers, and coals. *Chem. Eur. J.* **9**, 4764–4770 (2003).
- Rosi, N. L. *et al.* Hydrogen storage in microporous metal-organic frameworks. *Science* **300**, 1127–1129 (2003).
- Mao, W. L. *et al.* Hydrogen clusters in clathrate hydrate. *Science* **297**, 2247–2249 (2002).
- Patchkovskii, S. & Tse, J. S. Thermodynamic stability of hydrogen clathrates. *Proc. Natl Acad. Sci. USA* **100**, 14645–14650 (2003).
- Mao, W. L. & Mao, H. Hydrogen storage in molecular compounds. *Proc. Natl Acad. Sci. USA* **101**, 708–710 (2004).
- Florusse, L. J. *et al.* Stable low-pressure hydrogen clusters stored in a binary clathrate hydrate. *Science* **306**, 469–471 (2004).
- Jeffrey, G. A. in *Comprehensive Supramolecular Chemistry* Vol. 6 (eds MacNicol, D. D., Toda, F. & Bishop, R.) 757–788 (Pergamon, Oxford, 1996).
- Nakamoto, K. *Infrared and Raman Spectra of Inorganic and Coordination Compounds* 4th edn (Wiley, New York, 1986).
- Nakahara, J. *et al.* C. C. Raman spectra of natural clathrates in deep ice cores. *Phil. Mag. B* **3**, 421–430 (1988).
- Lokshin, K. A. *et al.* Structure and dynamics of hydrogen molecules in the novel clathrate hydrate by high pressure neutron diffraction. *Phys. Rev. Lett.* **93**, 125503 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was supported by the Korea Research Foundation and the Brain Korea 21 Project.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.L. (h_lee@kaist.ac.kr) or J.A.R. (John.Ripmeester@nrc-cnrc.gc.ca).

Water content in the transition zone from electrical conductivity of wadsleyite and ringwoodite

Xiaoge Huang^{1,2}, Yousheng Xu² & Shun-ichiro Karato²

¹Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China

²Department of Geology and Geophysics, Yale University, New Haven, Connecticut 06511, USA

The distribution of water in the Earth's interior reflects the way in which the Earth has evolved, and has an important influence on its material properties. Minerals in the transition zone of the Earth's mantle (from ~ 410 to ~ 660 km depth) have large water solubility^{1–3}, and hence it is thought that the transition zone might act as a water reservoir. When the water content of the transition zone exceeds a critical value, upwelling flow might result in partial melting at ~ 410 km, which would affect the distribution of certain elements in the Earth⁴. However, the amount of water in the transition zone has remained unknown. Here we determined the effects of water and temperature on the electrical conductivity of the minerals wadsleyite and ringwoodite to infer the water content of the transition zone. We find that the electrical conductivity of these minerals depends strongly on water content but only weakly on temperature. By comparing these results with geophysically inferred conductivity^{5–7}, we infer that the water content in the mantle transition zone varies regionally, but that its value in the Pacific is estimated to be ~ 0.1 – 0.2 wt%. These values significantly exceed the estimated critical water content in the upper mantle^{3,8,9}, suggesting that partial melting may indeed occur at ~ 410 km depth, at least in this region.

Water (hydrogen) plays an important role in a number of geophysical or geochemical processes. In particular, melting behaviour of silicate rocks^{10–12}, plastic deformation^{13–15}, electrical conductivity¹⁶ and diffusion¹⁷ are known to be affected significantly by the presence of water. Consequently, determining the distribution of water in the Earth's mantle is an important issue in solid Earth geophysics and geochemistry. Experimental studies have shown that the mantle transition zone can be a large reservoir for water, and a water-enriched transition zone may be important in geochemical cycling⁴, but the exact amount of water in the transition zone is poorly constrained.

A promising way to infer the water content in the actual Earth is to compare some geophysical observations with experimental data on the effects of water¹⁸. Here we investigated the effects of water on the electrical conductivity of wadsleyite and ringwoodite, and by comparing the results with geophysically determined electrical conductivity in the transition zone, we estimated the water content in the transition zone.

Wadsleyite and ringwoodite samples with a grain size of 5–10 μm were prepared at pressure $P = 14$ – 16 GPa and temperature

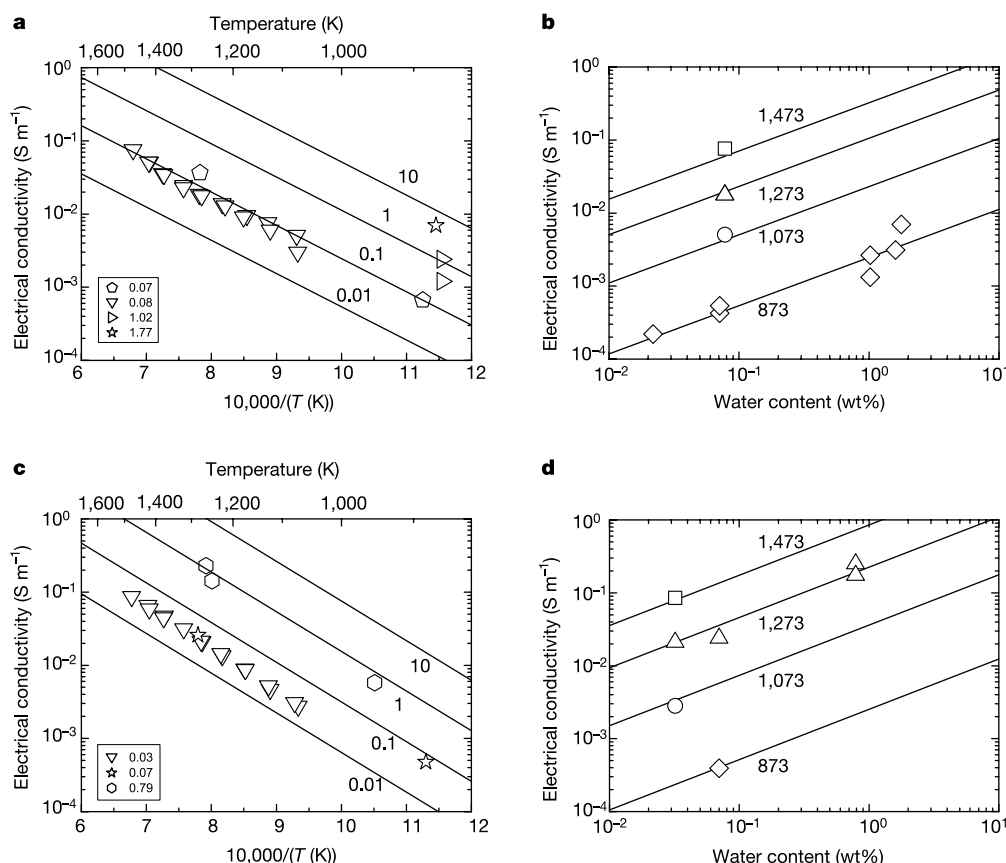


Figure 1 Measured electrical conductivity. **a**, Electrical conductivity of wadsleyite as a function of temperature for several water contents; **b**, electrical conductivity of wadsleyite as a function of water content for several temperatures; **c**, electrical conductivity of ringwoodite as a function of temperature for various water contents; **d**, electrical conductivity of ringwoodite as a function of water content for various temperatures. All

data are at pressures $P = 14\text{--}16$ GPa. The data corresponding to inverted triangles in **a** and **c** are taken from figure 3B in ref. 20 and their water contents obtained from the recovered samples are also included in the data fitting. The lines in **b** and **d** are calculated from equation (1) and Table 1 for wadsleyite and ringwoodite, respectively. In **a** and **c**, numbers present water content in wt%; in **b** and **d**, numbers present temperatures in K.

$T = 1,273\text{--}1,473$ K with controlled water contents using a Kawai-type multianvil apparatus at Yale University. The starting material was San Carlos olivine, and $\sim 5\%$ orthopyroxene was added to control the oxide activity to the silica-rich end. The oxygen fugacity was buffered by the Mo/MoO₂ solid-state reaction. Synthesis was performed either under water-free conditions or with a certain amount of free water. The concentration of hydrogen in samples was determined by Fourier-transform infrared (FT-IR) spectroscopy both before and after each measurement of electrical conductivity, using a $\sim 30\text{-}\mu\text{m}$ -thick sample in crack-free regions of $\sim 40\text{-}\mu\text{m} \times 40\text{-}\mu\text{m}$. The Paterson calibration¹⁹ was used to calculate hydrogen content from infrared absorption.

A thin disk of a sample was cut and the electrical conductivity was measured at $P = 14\text{--}16$ GPa and $T = 773\text{--}1,273$ K in the frequency range $10^2\text{--}10^6$ Hz using the technique developed in ref. 20. The pressure was estimated from the load-pressure calibration using several pressure standards, and the temperature was measured by a W5%Re–W26%Re thermocouple. After pressurization, temperature was increased to a desired value at a rate of ~ 50 K min^{−1} below 873 K and ~ 200 K min^{−1} above 873 K. The rapid temperature ramp at high temperature was chosen to minimize water loss. All recovered samples were examined using a micro-Raman spectrometer for phase identification. In most runs, the water content changed during a run: in water-poor samples, the water content increased, and in water-rich samples, the water content decreased. The change in water content is typically small ($\sim 30\%$ or less), and we estimate that the overall uncertainty of hydrogen content for each measurement is better than $\sim 20\%$. We determined the

frequency-dependent impedance that has both real and complex components. The absolute value of impedance approaches an asymptotic value at low frequencies. This asymptotic value is used to calculate the conductivity of a sample. The conductivity values corresponding to these asymptotic values correspond to those determined by geophysical methods. The results are shown in Fig. 1.

We analyse the results assuming the following relationship:

$$\sigma = AC_W^r \exp\left(-\frac{H^*}{RT}\right) \quad (1)$$

where σ is electrical conductivity, A and r are constants, C_W is water content, H^* is activation enthalpy, R is the gas constant and T is temperature. The influence of grain size on electrical conductivity was not investigated systematically in this study. However, previous studies on Fe-bearing olivine showed no appreciable effect of grain size. Therefore we do not consider the grain-size effect here. The parameters for wadsleyite and ringwoodite are summarized in Table 1. Assuming that water (hydrogen) affects the conductivity through its influence on defect concentration, the results can be

Table 1 Parameter values

Mineral	A (S m ^{−1})	r	H^* (kJ mol ^{−1})
Wadsleyite	380 (+170, −120)	0.66 (±0.05)	88 (±3)
Ringwoodite	4070 (+1,050, −840)	0.69 (±0.03)	104 (±2)

Parameters were derived from fitting the equation $\sigma = AC_W^r \exp(-H^*/RT)$ to the experimental conductivity data. Numbers in parentheses are the errors (one standard deviation).

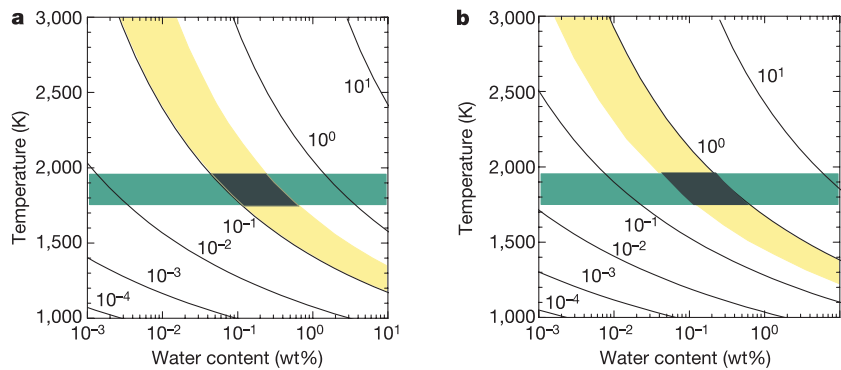


Figure 2 The influence of temperature and water content on electrical conductivity. **a**, Wadsleyite; **b**, ringwoodite. The experimental results summarized in Table 1 are corrected for oxygen fugacity effects (see text for details). The numbers on the curves represent electrical conductivity (in S m^{-1}). The light green shaded areas represent temperatures of $1,850 \pm 100 \text{ K}$; the yellow shaded areas represent the electrical

conductivity ranges in **a** the upper, and **b** the lower, transition zone in the Pacific⁴; the dark green areas represent the range of temperature and water content that is consistent with the observations. The water content in the transition zone in the Pacific is $\sim 0.1\text{--}0.2 \text{ wt\%}$ inferred from both wadsleyite and ringwoodite.

interpreted by using the Nernst–Einstein relation, $\sigma = \sum n_j q_j \mu_j$ (where n_j is the concentration of the j th type of charge carrier, q_j is the effective charge, and μ_j is its mobility). The defect chemistry of minerals under hydrous conditions shows that the concentration of hydrogen-related defects $[X]$ depends on the chemical environment as:

$$[X] \propto f_{\text{H}_2\text{O}}^p f_{\text{O}_2}^q a_{\text{MeO}}^s \tag{2}$$

where p , q and s are constants that depend on the type of defect, $f_{\text{H}_2\text{O}}$ is water fugacity, f_{O_2} oxygen fugacity and a_{MeO} the activity of metal oxide (such as $(\text{Mg,Fe})\text{O}$). In particular, a model of water (hydrogen) dissolution in olivine, wadsleyite and ringwoodite shows that water is dissolved in these minerals mostly as a neutral defect $(2\text{H})_{\text{M}}^{\times}$ (two protons trapped at an M-site)³, and the water solubility is related to the chemical environment as:

$$C_{\text{W}} \propto f_{\text{H}_2\text{O}} f_{\text{O}_2}^{-1/8} a_{\text{MeO}}^{-1} \tag{3}$$

In our experiments, both f_{O_2} and a_{MeO} were fixed and $f_{\text{H}_2\text{O}}$ was different for different samples. Consequently the value of r in equation (1) must be identical to the value of p in equation (2) for the defect that carries most of the electric current. The values of p , q and s for typical defects in wadsleyite and ringwoodite are summarized in Table 2. A comparison of the experimental results with this table suggests that the charge-carrying species in wadsleyite and ringwoodite under the present experimental condition is the free proton and not the most abundant defect, that is, hydrogen trapped at an M-site. Given this model, the electrical conductivity is predicted to depend on chemical environment as:

$$\sigma \propto f_{\text{H}_2\text{O}}^{3/8} f_{\text{O}_2}^{-1/8} a_{\text{MeO}}^{-1} \exp\left(-\frac{H^*}{RT}\right) \tag{4}$$

The oxide activity in our experiments is controlled by the presence of excess $(\text{Mg,Fe})\text{SiO}_3$ which is probably the same in the Earth. The

oxygen fugacity in our experiments is controlled by the Mo/MoO_2 buffer. Oxygen fugacity in the Earth's transition zone is unknown, and therefore we consider a range of oxygen fugacity (from the value similar to that of the upper mantle, that is, fayalite–magnetite–quartz buffer, to Mo/MoO_2 buffer). The correction for oxygen fugacity is made using:

$$\sigma_{\text{Earth}}(T, C_{\text{W}}) = \left(\frac{f_{\text{O}_2, \text{Earth}}}{f_{\text{O}_2, \text{Lab}}}\right)^q \sigma_{\text{Lab}}(T, C_{\text{W}}) \tag{5}$$

where σ_{Earth} is the electrical conductivity of the Earth, σ_{Lab} is the electrical conductivity of wadsleyite and ringwoodite, $f_{\text{O}_2, \text{Lab}}$ is the oxygen fugacity corrections to the Mo/MoO_2 buffer, $f_{\text{O}_2, \text{Earth}}$ is the oxygen fugacity in the Earth, and q is a constant ($-1/8$). We assumed that the f_{O_2} in the mantle is close to that of the Ni/NiO buffer²¹. This will reduce the conductivity by a factor of ~ 2 ($\sigma_{\text{Earth}} \approx 0.5 \sigma_{\text{Lab}}$). Because the pressure effect on electrical conductivity is usually small²² and the variation of pressure in the transition zone is small, the pressure effect is ignored.

Figure 2 shows the trade-off between the effects of water content

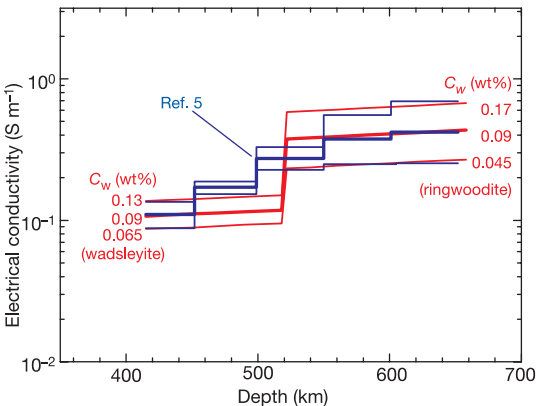


Figure 3 A comparison of laboratory data on electrical conductivity as a function of water content with the geophysically inferred electrical conductivity in the mantle transition zone in the Pacific⁵. C_{W} indicates water content in wt%. The blue lines represent the geophysically inferred conductivity in the Pacific and the thin blue lines represent the upper and lower bounds. The red lines correspond to those calculated from electrical conductivity of wadsleyite for 410–520 km and from electrical conductivity of ringwoodite for 520–660 km for a range of water contents (the effects of oxygen fugacity are corrected). The adiabatic temperatures ($1,825\text{--}1,900 \text{ K}$ in the transition zone) were used in the calculation. The results show a hydrogen content of $\sim 0.1\text{--}0.2 \text{ wt\%}$ in the transition zone in this region.

Table 2 Point defect concentrations							
Defect	p	q	s	Defect	p	q	s
$[\text{H}_{\text{M}}^{\times}]$	$\frac{1}{2}$	$\frac{1}{8}$	$-\frac{1}{2}$	$[(2\text{H})_{\text{M}}^{\times}]$	1	0	-1
$[\text{H}^{\times}]$	$\frac{1}{2}$	$-\frac{1}{8}$	$-\frac{1}{2}$	$[\text{H}_{\text{Si}}^{\times}]$	$-\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
$[(2\text{H})_{\text{Si}}^{\times}]$	$\frac{1}{2}$	$\frac{1}{4}$	2	$[(3\text{H})_{\text{Si}}^{\times}]$	$\frac{5}{4}$	$\frac{1}{8}$	$\frac{3}{2}$
$[(4\text{H})_{\text{Si}}^{\times}]$	2	0	1	$[(\text{OH})_{\text{Si}}^{\times}]$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$

Table shows dependence of concentrations of point defects in wadsleyite and ringwoodite on chemical environment: $[X] \propto f_{\text{H}_2\text{O}}^p f_{\text{O}_2}^q a_{\text{MeO}}^s$. The charge neutrality condition of $[\text{H}_{\text{M}}^{\times}] = [\text{Fe}_{\text{M}}^{\times}]$ is assumed. The Kröger–Vink notation of point defects is used: subscript indicates the site of the defect, prime indicates effective negative charge, superscript dot indicates effective positive charge and superscript cross indicates neutral charge.

and temperature on the electrical conductivity of wadsleyite and ringwoodite. Owing to the small activation enthalpy, the temperature effect is relatively small: a variation in temperature of ~ 200 K results in a change in conductivity of a factor of ~ 2 . In contrast, the expected water contents in the transition zone range from almost dry ($\sim 10^{-4}$ wt%) to water-saturated (~ 3 wt%), which corresponds to the variation of conductivity by a factor of $\sim 10^3$. Therefore water content is well-constrained by the conductivity data.

The electrical conductivity in the transition zone ranges from $\sim 10^{-2}$ S m $^{-1}$ to ~ 1 S m $^{-1}$ (refs 5–7). This corresponds to water contents of ~ 0.001 to ~ 0.4 wt%. For the upper mantle of the North Pacific Ocean, where a detailed inversion was made using mineral physics constraints⁵, the conductivity in the transition zone is $\sim 10^{-1}$ to $\sim 5 \times 10^{-1}$ S m $^{-1}$ and the corresponding water content in the transition zone is estimated to be ~ 0.1 – 0.2 wt% for the temperature range of 1,825–1,900 K (Fig. 3). The estimated water content in the north Pacific transition zone (~ 0.1 – 0.2 wt%) is significantly higher than the estimated water content in the upper mantle^{8,9} and probably exceeds a critical concentration for partial melting. This suggests that there is a marked layering in water content in the Earth's mantle. This is consistent with a recent model of material circulation⁴ involving the separation of the circulation of incompatible elements from that of major elements near 410 km, but is not consistent with a model involving vertical circulation of hydrogen together with major minerals. However, some regional variation in electrical conductivity was also reported^{5,6}, suggesting a large regional variation in hydrogen content (and temperature) in the transition zone. □

Received 16 September 2004; accepted 31 January 2005; doi:10.1038/nature03426.

1. Smyth, J. R. β -Mg₂SiO₄: A potential host for water in the mantle? *Am. Mineral.* **72**, 1051–1055 (1987).
2. Kawamoto, T., Hervig, R. L. & Holloway, J. R. Experimental evidence for a hydrous transition zone in the early Earth's mantle. *Earth Planet. Sci. Lett.* **142**, 587–592 (1996).
3. Kohlstedt, D. L., Keppler, H. & Rubie, D. C. The solubility of water in α , β and γ phases of (Mg,Fe)₂SiO₄. *Contrib. Mineral. Petrol.* **123**, 345–357 (1996).
4. Bercowski, D. & Karato, S. Whole-mantle convection and the transition-zone water filter. *Nature* **425**, 39–44 (2003).
5. Utada, H., Koyama, T., Shimizu, H. & Chave, A. D. A semi-global reference model for electrical conductivity in the mid-mantle beneath the north Pacific region. *Geophys. Res. Lett.* **30**, 1194–1198 (2003).
6. Ichiki, M. *et al.* Upper mantle conductivity structure of the back-arc region beneath northeastern China. *Geophys. Res. Lett.* **28**, 3773–3776 (2001).
7. Tarits, P., Hautot, S. & Perrier, F. Water in the mantle: Results from electrical conductivity beneath the French Alps. *Geophys. Res. Lett.* **31**, L06612, doi:10.1029/2003GL019277 (2004).
8. Bell, D. R. & Rossman, G. R. Water in Earth's mantle—the role of nominally anhydrous minerals. *Science* **255**, 1391–1397 (1992).
9. Hirth, G. & Kohlstedt, D. L. Water in oceanic upper mantle—implications for rheology, melt extraction and the evolution of lithosphere. *Earth Planet. Sci. Lett.* **144**, 93–108 (1996).
10. Kushiro, I. *et al.* Melting of a peridotite nodule at high pressures and high water pressures. *J. Geophys. Res.* **73**, 6023–6029 (1968).
11. Hirose, K. & Kushiro, I. Partial melting of dry peridotites at high pressures: Determination of compositions of melts segregated from peridotite using aggregates of diamond. *Earth Planet. Sci. Lett.* **114**, 477–489 (1993).
12. Grove, T. L. *et al.* Fractional crystallization and mantle melting controls on calc-alkaline differentiation trends. *Contrib. Mineral. Petrol.* **145**, 515–533 (2003).
13. Karato, S., Paterson, M. S. & Fitz Gerald, J. D. Rheology of synthetic olivine aggregates—implication of grain-size and water. *J. Geophys. Res.* **91**, 8151–8176 (1986).
14. Mei, S. & Kohlstedt, D. L. Influence of water on plastic deformation of olivine aggregates: 1. Diffusion creep regime. *J. Geophys. Res.* **105**, 21457–21469 (2000).
15. Karato, S. & Jung, H. Effects of pressure on high-temperature dislocation creep in olivine. *Phil. Mag.* **83**, 401–414 (2003).
16. Karato, S. The role of hydrogen in the electrical conductivity of the upper mantle. *Nature* **347**, 272–273 (1990).
17. Farver, J. R. & Yund, R. A. Oxygen fugacity in quartz: dependence on temperature and water fugacity. *Chem. Geol.* **90**, 55–70 (1991).
18. Karato, S. in *Inside the Subduction Factory* (ed. Eiler, J.) 138–152 (Am. Geophys. Union, Washington DC, 2003).
19. Paterson, M. S. The determination of hydroxyl by infrared absorption in quartz, silicate glasses and similar materials. *Bull. Mineral.* **105**, 20–29 (1982).
20. Xu, Y., Poe, B. T., Shankland, T. J. & Rubie, D. C. Electrical conductivity of olivine, wadsleyite, and ringwoodite under upper-mantle conditions. *Science* **280**, 1415–1418 (1998).
21. Wood, B. J., Bryndzia, L. T. & Johnson, K. E. Mantle oxidation state and its relationship to tectonic environment and fluid speciation. *Science* **248**, 337–345 (1990).
22. Xu, Y., Shankland, T. J. & Duba, A. G. Pressure effect on electrical conductivity of mantle olivine. *Phys. Earth Planet. Inter.* **118**, 149–161 (2000).

Acknowledgements Z. Jiang and Y. Nishihara provided the technical assistance that made this research possible. This work was supported by the NSF of China and the NSF of the United States.

Authors' contributions S.-I.K. supervised the whole project. The experimental measurements of electrical conductivity were made by X.H. in collaboration with Y.X., and the theoretical interpretation of the results and the geophysical applications were made by S.-I.K. together with Y.X.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.-I.K. (shun-ichiro.karato@yale.edu).

Particle size and energetics of gouge from earthquake rupture zones

Brent Wilson¹, Thomas Dewers¹, Ze'ev Reches¹ & James Brune²

¹School of Geology and Geophysics, University of Oklahoma, Norman, Oklahoma 73019, USA

²Department of Geological Sciences, University of Nevada, Reno, Nevada 89557, USA

Grain size reduction and gouge formation are found to be ubiquitous in brittle faults at all scales^{1–4}, and most slip along mature faults is observed to have been localized within gouge zones^{5,6}. This fine-grain gouge is thought to control earthquake instability^{3,6–8}, and thus understanding its properties is central to an understanding of the earthquake process^{7,9}. Here we show that gouge from the San Andreas fault, California, with ~ 160 km slip, and the rupture zone of a recent earthquake in a South African mine with only ~ 0.4 m slip, display similar characteristics, in that ultrafine grains approach the nanometre scale, gouge surface areas approach $80 \text{ m}^2 \text{ g}^{-1}$, and grain size distribution is non-fractal. These observations challenge the common perception that gouge texture is fractal^{10,11} and that gouge surface energy is a negligible contributor to the earthquake energy budget^{3,9,12}. We propose that the observed fine-grain gouge is not related to quasi-static cumulative slip, but is instead formed by dynamic rock pulverization during the propagation of a single earthquake.

Gouge formation is commonly attributed to wear and attrition of sliding surfaces^{1–3,5,7} or to implosive loading¹³ during earthquakes. Models of wear and attrition¹⁰ predict that progressive shear within gouge zones should lead to grain size reduction due to amplified stresses at grain contacts, and to fractal particle size distributions¹⁰ of the gouge; these predictions are supported by field and experimental observations^{1,2,10,11} with a few exceptions².

However, there are two main pitfalls in standard textural analysis of fine-grained gouge¹⁴. First, optical microscopic analysis and mechanical sieving are restricted to grains a few micrometres in diameter or larger. Second, measurements of the size distribution and surface area of fine-grained materials are susceptible to erroneous results due to the aggregation or agglomeration of extremely fine particles¹⁵. For example, our preliminary measurements of the particle size distribution (PSD) of fault gouge with a laser particle-size analyser reveal profound time drift: the mean grain size of pulverized granite from the San Andreas fault zone (see below) decreased by $26 \pm 13\%$ (s.d.) during 0.5 h of continuous measurement (on the basis of 145 samples¹⁶). This time drift reflects disaggregation into primary grains, as is known for clay mineral aggregates¹⁷. Questions therefore remain open with regard to the size, PSD and surface area of primary particles in fault gouge. To determine these properties accurately we developed a new pro-

cedure of continuous measurements (see Methods). The range of measured particles, down to grains tens of nanometres in size, are also observed directly by scanning electron microscopy.

Pristine gouge is difficult to find. Chemical alteration and lithification alter the texture in exhumed fault zones¹⁸, gouge found at surface rupture zones cannot reflect mechanical conditions at depth, and rupture zones are usually inaccessible at earthquake focal depths. We chose two fault systems that partly remove these limitations. The first is the San Andreas fault, which is a system 1,200 km long that accommodates hundreds of kilometres of right-lateral slip. The better-known exposures of deep parts of the San Andreas fault are inactive segments in southern California that were uplifted from 2–4 km depth^{5,6,18}. Our study focuses on San Andreas gouge exposed in the Tejon Pass region¹⁹ (~100 km north of Los Angeles), alongside the rupture zone of the $M = 8.0$ earthquake in 1857 (Fig. 1a). Here, three sub-parallel fault segments bound a gouge zone 70–100 m wide composed of pervasively pulverized granite¹⁹, distinctly different in PSD and morphology from weathered granites outside the fault zone¹⁹. On the basis of current uplift rates along the San Andreas system²⁰, we estimate that the Tejon Pass region has experienced as much as 4–6 km of uplift since the Pliocene epoch, and that the observed gouge formed at this estimated depth.

South African gold mines, 1.0–3.5 km deep, are shaken daily by thousands of earthquakes. The mines provide access to rupture zones of these earthquakes at focal depth, a unique opportunity for earthquake investigations^{4,21,22}. Gouge is analysed from the rupture zone of the 1997 $M = 3.7$ earthquake in the Hartbeestfontein gold mine (~120 km southwest of Johannesburg, South Africa). This event occurred within unfaulted quartzitic layers at ~2 km depth and produced a new fault (the Bosman fault) later uncovered by mining operations. The Bosman fault zone is at least 5 m wide and 100 m long with 0.37 m of maximum dip-slip displacement, and contains four to six large subparallel segments with hundreds of secondary small fractures (Fig. 1b). Fractures are filled with white gouge that is commonly observed in brittle failure zones of quartzite in mines²³. The gouge of the newborn Bosman fault formed from a single earthquake and was not affected by cumulative slip of multiple seismic events.

The present study includes PSD measurements of ~250 gouge samples from both faults; 155 samples were measured for 0.5 h or more, with eight samples being measured for 45–190 h (see Methods). The results are shown by the following: first, the initial and final PSD (after 72 and 190 h) of the gouge (Fig. 2a); second, the initial and final cumulative frequencies of grain numbers (Fig. 2b); and third, the power-law time decay of mean grain size and the associated increase in surface area of gouge from both faults (Fig. 3). Because the results for the long-duration measurements are generally similar, the figures show only representative samples.

We determined geometric surface area from the PSD of the gouge first by assuming smooth, spherical grains. Independently, the surface areas of six untreated samples of San Andreas gouge were measured with the Barrett–Emmett–Teller (BET) N_2 adsorption technique¹⁵. To estimate grain roughness, we calculated $\lambda = (\text{BET surface area})/(\text{geometric surface area})$; the λ value, 6.6 ± 1.5 , for these samples is similar to a known range of λ of 5.5–22 (ref. 24). With the use of $\lambda = 6.6$, the calculated geometric surface area is converted to equivalent BET surface area (Fig. 3b), revealing surface area values as high as $80 \text{ m}^2 \text{ g}^{-1}$. This surface area is probably a conservative estimate because grain disaggregation does not reach asymptotic values by the end of the long runs (Fig. 3). Scanning electron microscope images of gouge aggregates can also be used to estimate surface area (in $\text{m}^2 \text{ g}^{-1}$) of roughly cubic fragments as $\sim 2/L$, where L is the grain size in micrometres. Thus, fragments of $0.02 \text{ } \mu\text{m}$ of aggregated gouge (Fig. 4) correspond to a surface area of $\sim 100 \text{ m}^2 \text{ g}^{-1}$.

Inspection of Figs 2 and 3 reveals the following three properties.

First, gouge PSD from both faults varies profoundly with run time, changing from an initial range of $0.04\text{--}200 \text{ } \mu\text{m}$ to less than $1 \text{ } \mu\text{m}$ after 72–190 h (Fig. 2a). The mean grain size does not attain asymptotic values (Fig. 3a), indicating that primary grains, which are tens of nanometres in size, might stay aggregated even after long runs. Accordingly, the apparent surface area of gouge grains increases with time and the power-law trend indicates a surface area of $100 \text{ m}^2 \text{ g}^{-1}$ after 300–500 h (Fig. 3b).

Second, gouge samples from the two faults display similar behaviour and final PSD even though the two faults are from strikingly different settings and magnitude.

Third, the grain number–size distribution of disaggregated gouge does not fit a systematic fractal distribution (Fig. 2b) because different slopes (corresponding to different fractal dimensions) are attained with progressive disaggregation. It is therefore inferred that the commonly observed fractal PSD^{1,2,10,11} reflects aggregated gouge grains in a restricted range of coarser grain sizes.

These results permit a better estimation of the earthquake energy balance. The gouge surface area of $\sim 80 \text{ m}^2 \text{ g}^{-1}$ (Fig. 3b) corre-

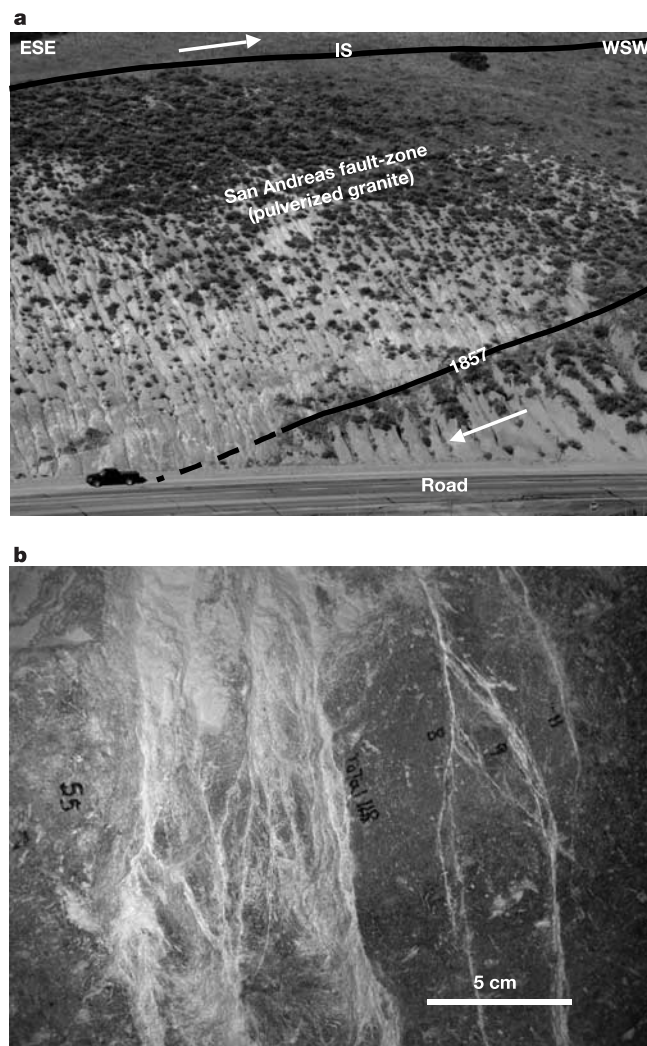


Figure 1 Field setting of the investigated faults. **a**, The San Andreas fault-zone at Tejon Pass, California. The car is shown for scale. IS, inactive fault segment; 1857, slipping segment of the 1857 earthquake. Note the gouge zone 70–100 m wide with badland morphology of the pulverized granite. This exposure and its extensions were mapped, and hundreds of gouge samples were collected. **b**, Close-up view of the fractured zone of the Bosman fault (text). The white striations are coalescing fractures filled with gouge powder (rock flour) that cut across the solid, dark quartzite.

sponds to a surface energy of 0.2–0.36 MJ per m² of the fault surface for a gouge zone 1 mm thick (for a specific surface energy of quartz of 1–1.8 J m⁻²). The Bosman fault earthquake generated tens of subparallel fractures that are each ~1 mm thick and filled with gouge (Fig. 1b); summation of the surface energy, marked here by U_s , of 10–30 fractures yields $U_s \approx 3$ –10 MJ m⁻². This value is roughly equal to the frictional energy calculated for the Bosman earthquake, $U_f = \tau d \approx 8$ –12 MJ m⁻² (τ is the estimated shear stress and d is the measured slip) and it is also similar to the energy release rate of earthquakes of 0.5–5 MJ m⁻² (ref. 25) to 1–100 MJ m⁻² (ref. 26). These energy relations suggest that the surface energy consumed by new gouge that forms during an earthquake can account for 50% or more of earthquake energy.

Finally, we conclude that the observed fine-grained gouge did not form by quasi-static wear and attrition, but rather formed by dynamic rock pulverization during earthquake propagation. This conclusion is based on two outcomes of the analysis. First, the similarity between the PSD of San Andreas gouge with ~160 km of slip at Tejon Pass, and the PSD of the Bosman fault gouge formed by one earthquake with 0.37 m slip, clearly indicates a lack of correlation between slip amount and grain size. Further, because a single earthquake generates the fine-grain gouge, it seems that the governing parameters of gouge formation are earthquake processes and not the cumulative slip attrition. Second, there is an apparent difference between energy consumption by gouge formation during

an earthquake (for example the Bosman fault) and during wear in quasi-static experiments²⁷. In the former, gouge formation consumed ~50% of the earthquake energy (see above), whereas in the latter, gouge formation consumed only 0.2–0.3% of the supplied mechanical energy (we calculate from ref. 27 that $U_f \approx 7$ MJ m⁻² and $U_s \approx 1.5 \times 10^4$ J m⁻², by assuming that their experimental gouge has the same surface area as that in the Bosman fault). An earthquake can pulverize rocks by a few mechanisms: one is the sequence of fault-normal unloading followed by implosive loading during the earthquake passage¹³; another results from the deformation conditions at the tip of the earthquake rupture²⁸. We calculated a maximum tension of ~5 GPa and dilation rates of $\sim 3 \times 10^5$ s⁻¹ at the tip of a mode II fracture propagating at 80% of the shear wave velocity²⁸; these intense conditions are comparable to shock loading²⁹ and could pulverize rocks into gouge.

We now apply the proposed dynamic rock pulverization to the San Andreas fault. If each earthquake in the studied area (Fig. 1a) generated a gouge zone 10 mm thick with ~ 80 m² g⁻¹ (Fig. 3b), the corresponding surface energy would be 2.0–3.6 MJ m⁻² for each

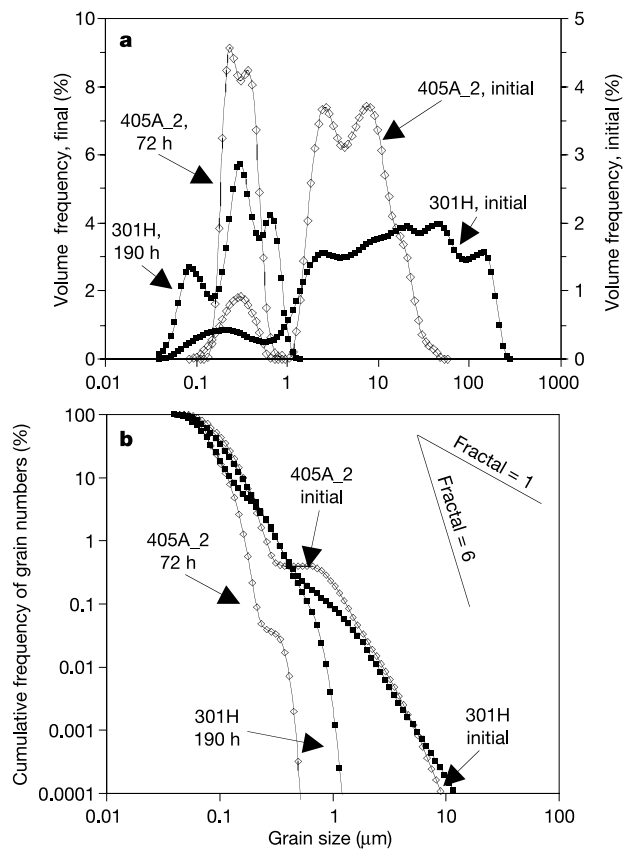


Figure 2 Particle size distribution (PSD) of two representative gouge samples measured for extended periods in a laser PSD analyser (see Methods). Samples: 301H, San Andreas fault (solid squares) and 405A_2, Bosman fault, South Africa (open diamonds). **a**, Grain size frequencies by volume. Each sample is displayed by two curves, one for the initial stage (right axis) and one for the final stage (left axis). A small fraction of the initial stage is smaller than 1 μ m, whereas almost all grains of the final stage are smaller than 1 μ m. **b**, Cumulative frequencies of grain numbers. A self-similar population should appear linear here, and fractal slopes of 1 and 6 (marked) seem to bound most of the curves.

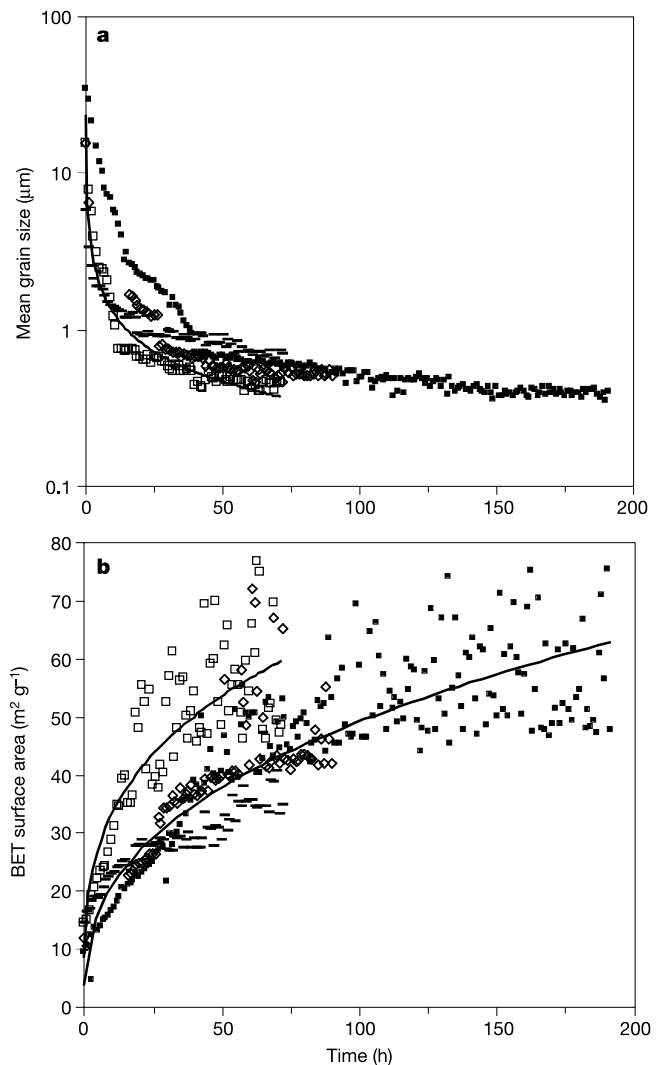


Figure 3 Time drift for four gouge samples measured for 72–190 h. **a**, Mean grain size; **b**, surface area. The plotted equivalent BET surface areas were calculated from the spherical grain area and the roughness parameter λ (text). Samples X-15 (open squares) and 301H (filled squares) are from the San Andreas fault; samples 405A_2 (open diamonds) and 405A_4 (filled oblongs) are from the Bosman fault, South Africa; four additional long-term runs with similar results are not plotted for clarity. Solid lines are power-law fits to representative samples.

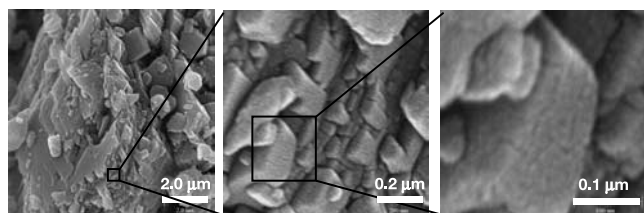


Figure 4 Scanning electron microscope images of untreated gouge from San Andreas gouge, with an order-of-magnitude resolution increase from left to right. The gouge samples were mounted onto colloidal graphite coating covering upper portions of platinum trays, and sputter-coated with gold–palladium under vacuum. Note the particles of 0.02–0.05 μm in the middle and right frames.

event. In such a case, the observed gouge zone 70–100 m wide was formed by 7,000–10,000 earthquakes, which is in agreement with estimates of recurrence intervals. Although the present observations on gouge energetics are in accord with some studies³⁰, they contradict common thought that gouge surface energy is a negligible component of earthquake energy balance^{3,9}. If our conclusions are valid in general, they could explain, for example, the heat flow anomaly of the San Andreas fault system⁶. □

Methods

We employ a Beckman Coulter LS230 laser diffraction particle size analyser. Its 750-nm laser source and proprietary polarization intensity differential scattering (PIDS) technology provide detection limits of 0.04–2,000 μm . Spectrometry by laser diffraction does not discern between primary particles and agglomerates and is therefore sensitive to the degree of agglomeration inside the analyser^{14,17}. Disaggregation is a time-dependent process that can occur over the course of days in silicate mineral suspensions¹⁷. Accordingly, gouge PSD measurements lasting up to 190 h were taken, during which progressive disaggregation could be discerned (Figs 2, 3). Initial sonication accelerated disaggregation but had no noticeable influence on the final PSD. Ultrafine particles might reaggregate during the analysis¹⁴, as indicated by the increased scatter in surface area at long times (Fig. 3b). Power-law disaggregation (Fig. 3) and recurring agglomeration/disaggregation during analysis indicates that PSD and surface area results are conservative estimates of primary gouge particle size and area produced by the seismic slip. Ultrafine particles could also have been lost as a result of Ostwald ripening and volatilization during sampling and handling.

The collected gouge samples were sealed at the site and stored in plastic bags. For the PSD measurements, tens of micrograms of sample were added to 25 ml of an aqueous surfactant solution (usually 1% analytical reagent grade sodium metaphosphate prepared with doubly distilled water) or methanol and then subjected for 30 min to a low-energy sonic bath. After an additional 30 min this slurry was added to the laser analyser containing 125 ml of the same solution. Measurements of the diffraction spectrum were performed with continuous circulation inside the analyser, and PIDS was used in all reported runs. Spectral analysis was performed with proprietary software using the Mie scattering model¹⁴, with constants for the complex refractive index plus wavelength dependence for quartz¹⁴ and an absorption coefficient of 0.01.

Received 4 December 2004; accepted 28 January 2005; doi:10.1038/nature03433.

1. Sammis, C. G., Osborne, R. H., Anderson, J. L., Banerdt, M. & White, P. Self-similar cataclasis in the formation of fault gouge. *Pure Appl. Geophys.* **124**, 53–78 (1986).
2. Marone, C. & Scholz, C. Particle-size distribution and microstructures within simulated fault gouge. *J. Struct. Geol.* **11**, 799–814 (1989).
3. Scholz, C. H. *The Mechanics of Earthquakes and Faulting* (Cambridge Univ. Press, London, 2002).
4. Dor, O., Reches, Z. & van Aswegen, G. in *Rockburst and Seismicity in Mines* Vol. 5 (eds van Aswegen, G., Durrheim, R. J. & Ortlepp, W. D.) 109–112 (South African Inst. of Mining and Metallurgy, Johannesburg, 2001).
5. Chester, F. M., Evans, J. P. & Biegel, R. L. Internal structure and weakening mechanisms of the San Andreas Fault. *J. Geophys. Res. Solid Earth* **98**, 771–786 (1993).
6. Zoback, M. D., Hickman, S. & Ellsworth, W. San Andreas Fault observatory at depth. (<http://www.icdp.gfzptsdam.de/html/sites/sanandreas/objectives/proposal.html>) (2002).
7. Ben-Zion, Y. & Sammis, C. G. Characterization of fault zones. *Pure Appl. Geophys.* **160**, 677–715 (2003).
8. Sleep, N. H. & Blanpied, M. L. Creep, compaction, and the weak rheology of major faults. *Nature* **359**, 687–692 (1992).
9. Olgaard, D. & Brace, W. The microstructure of gouge from a mining-induced seismic shear zone. *Int. J. Rock Mech. Mining Sci.* **20**, 11–19 (1983).
10. Steacy, S. J. & Sammis, C. G. An automaton for fractal patterns of fragmentation. *Nature* **353**, 250–252 (1991).
11. An, L. J. & Sammis, C. G. Particle-size distribution of cataclastic fault materials from southern California—a 3-d study. *Pure Appl. Geophys.* **143**, 203–227 (1994).
12. Kanamori, H. Mechanics of earthquakes. *Annu. Rev. Earth Planet. Sci.* **22**, 207–237 (1994).
13. Brune, J. N. Fault-normal dynamic unloading and loading: an explanation for 'nongouge' rock powder and lack of fault-parallel shear bands along the San Andreas Fault. *Eos* **8**, 47 (2001).
14. Xu, R. *Particle Characterization: Light Scattering Methods* (Kluwer Academic, Dordrecht, 2000).

15. Gregg, S. J. & Sing, K. S. W. *Adsorption, Surface Area and Porosity* (Academic, London, 1982).
16. Dewers, T. A., Wilson, B. & Reches, Z. Scaling particle size in fault gouge: Variable fractal dimension or non-fractal distribution? *Eos* **84**, NG12C-06 (2003).
17. Franco, F., Perez-Maqueada, L. A. & Perez-Rodriguez, J. L. The effect of ultrasound on the particle size and structural disorder of a well-ordered kaolinite. *J. Colloid Interface Sci.* **274**, 107–117 (2004).
18. Evans, J. P. & Chester, F. M. Fluid–rock interaction in faults of the San-Andreas System—Inferences from San-Gabriel fault rock geochemistry and microstructures. *J. Geophys. Res.* **100**, 13007–13020 (1995).
19. Wilson, B. *Meso- and Micro-structural Analysis of the San Andreas Fault at Tejon Pass, California*. Thesis, Univ. Oklahoma, Norman (2004).
20. Smith, B. & Dandwell, D. Coulomb stress accumulation along the San Andreas Fault system. *J. Geophys. Res.* **108**, 2296 (2003).
21. McGarr, A., Spottiswoode, S. M., Gay, N. C. & Ortlepp, W. D. Observations relevant to seismic driving stress, stress drop, and efficiency. *J. Geophys. Res.* **84**, 2251–2261 (1978).
22. Ogasawara, H., Yanagidani, Y. & Ando, M. (eds) *Seismogenic Process Monitoring* (Balkema, Rotterdam, 2002).
23. Ortlepp, W. D. *Rock Fracture and Rockbursts* (South Africa Institute of Mining and Metallurgy, Monograph series M9, 1997).
24. Hochella, M. F. Jr & Banfield, J. F. in *Chemical Weathering Rates of Silicate Minerals* (eds White, A. F. & Brantley, S. L.) 353–406 (Mineralogical Society of America, Washington DC, 1995).
25. Poliakov, A. N. B., Dmowska, R. & Rice, J. R. Dynamic shear rupture interactions with fault bends and off-axis secondary faulting. *J. Geophys. Res.* **107**, 2295 (2002).
26. Li, V. C. in *Fracture Mechanics of Rocks* (ed. Atkinson, B. K.) 351–428 (Academic, London, 1987).
27. Yund, R. A., Blanpied, M. L., Tullis, T. E. & Weeks, J. D. Amorphous material in high strain experimental fault gouges. *J. Geophys. Res.* **95**, 15589–15602 (1990).
28. Reches, Z. & Dewers, T. A. Gouge formation by dynamic pulverization during earthquakes. *Earth Planet. Sci. Lett.* (submitted).
29. Grady, D. E. & Kipp, D. E. Geometric statistics and dynamic fragmentation. *J. Appl. Phys.* **58**, 1210–1222 (1985).
30. Reches, Z. Mechanisms of slip nucleation during earthquakes. *Earth Planet. Sci. Lett.* **170**, 475–486 (1999).

Acknowledgements We thank the US National Science Foundation and the Southern California Earthquake Center for supporting this research.

Authors' contributions All authors contributed equally to this work.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Z.R. (reches@ou.edu).

New material of the earliest hominid from the Upper Miocene of Chad

Michel Brunet¹, Franck Guy^{1,2}, David Pilbeam², Daniel E. Lieberman², Andossa Likius³, Hassane T. Mackaye³, Marcia S. Ponce de León⁴, Christoph P. E. Zollikofer⁴ & Patrick Vignaud¹

¹Laboratoire de Géobiologie, Biochronologie et Paléontologie Humaine, CNRS UMR 6046, Faculté des Sciences, Université de Poitiers, 40 Avenue du Recteur Pineau, 86022 Poitiers Cedex, France

²Peabody Museum, Harvard University, 11 Divinity Avenue, Cambridge, Massachusetts 02138, USA

³Université de N'Djamena, BP 1117, N'Djamena, Tchad

⁴Anthropologisches Institut/MultiMedia Laboratorium, Universität Zürich-Irchel, Winterthurerstrasse 190, 8057 Zürich, Switzerland

Discoveries in Chad by the Mission Paléanthropologique Franco-Tchadienne have substantially changed our understanding of early human evolution in Africa^{1–3}. In particular, the TM 266 locality in the Toros-Menalla fossiliferous area yielded a nearly complete cranium (TM 266-01-60-1), a mandible, and several isolated teeth assigned to *Sahelanthropus tchadensis*³ and biochronologically dated to the late Miocene epoch (about 7 million years ago). Despite the relative completeness of the TM 266 cranium, there has been some controversy about its morphology and its status in the hominid clade^{4,5}. Here we describe new dental and mandibular specimens from three Toros-Menalla (Chad) fossiliferous localities (TM 247, TM 266 and TM 292) of

the same age⁶. This new material, including a lower canine consistent with a non-honing C/P₃ complex, post-canine teeth with primitive root morphology and intermediate radial enamel thickness, is attributed to *S. tchadensis*. It expands the hypodigm of the species and provides additional anatomical characters that confirm the morphological differences between *S. tchadensis* and African apes. *S. tchadensis* presents several key derived features consistent with its position in the hominid clade close to the last common ancestor of chimpanzees and humans.

The upper Miocene vertebrate localities from the Toros-Menalla fossiliferous area discovered by the Mission Paléoanthropologique Franco-Tchadienne in the Mega-Chad basin, are north of the 16th parallel, 150 km west of the Koro-Toro australopithecine localities^{1,2,7}. The faunal assemblage from TM 266 is found in the Anthracotheriid Unit, so named because it contains a very common, large anthracotheriid, *Libycosaurus petrochii*⁶. The mammalian fauna from the Anthracotheriid Unit, which includes a primitive suid, *Nyanzachoerus syrticus*, and a primitive loxodont elephant, contains more primitive taxa than the Lukeino fauna (Kenya, dating from 6 Myr ago)⁸ and is more similar to the fauna from the lower Nawata Formation of Lothagam (Kenya, 6.5–7.4 Myr ago)⁹. The Anthracotheriid Unit assemblage indicates a mosaic of landscapes⁶ probably resembling that of the present-day Okavango Delta (Botswana). Previous collecting in TM 266 uncovered a cranium, TM 266-01-60-1, as well as two mandibular fragments and several isolated teeth assigned to *Sahelanthropus tchadensis*³. Because of the age of this earliest hominid taxon (the term hominid is used here for convenience to denote all taxa that are closer to humans than chimpanzees, and does not connote any taxonomic scheme³; similarly, australopithecine is used as a generic term *sensu lato* to refer to all Pliocene hominid taxa that do not belong to the genera *Ardipithecus* and *Homo*), it is important to evaluate and expand the hypodigm to test hypotheses about its systematic relationships. Additional information that expands the *Sahelanthropus tchadensis* hypodigm comes from recent discoveries of new hominid material from TM 266 and from two new sites, TM 247 and TM 292, also in the Anthracotheriid Unit. These three sites are within a small area (0.73 km²). The new specimens (Table 1) consist of two lower jaws (Figs 1, 2) and the crown of a right P₃ (Fig. 3).

TM 292-02-01 (Fig. 1) is a partial mandible fragment lacking the left and right corpus posterior to M₂. The cortical bone is well preserved except in the antero-medial lower part of the symphyseal region, and in the alveolar process in the region of the incisors. The left I₂, C₁, M₁ and M₂ roots and right I₁–I₂, P₃ and M₁–M₂ roots are preserved. The crowns of the left M₁, M₂ and the right M₁ are partly preserved, and the crown of the left canine is well preserved (Fig. 1g, h). TM 247-01-02 (Fig. 2) is a fragmentary right mandibular corpus. All the roots are well preserved; the crowns of P₃–M₁ are partly preserved but are missing in M₂–M₃. The corpus of the TM 292-02-01 fragment is more gracile (maximum corpus breadth at M₁, perpendicular to corpus height, is 14.5 mm) than that of the previously discovered TM 266-02-154-1 specimen³ (maximum corpus breadth at M₁ is 20.0 mm) as well as the newly discovered TM 247-01-02 (corpus breadth at M₁ is 16.1 mm), although this is a minimum estimate because the cortical bone surface has been eroded on the buccal side of the corpus. TM 292-02-01 and TM 247-01-02 each have a single, large mental foramen located at mid-corpus below P₄. The anterior margin of the symphysis in

TM 292-02-01 is vertical (Fig. 1c, d) with the rather damaged inferior part sloping posteriorly. The planum alveolare of the symphysis is about 45° relative to the alveolar plane of the corpus. The inferior and superior transverse tori are weakly developed (superior is larger), and delimit a shallow genioglossal fossa with a large genioglossal foramen.

Among the mandibular teeth, only the lower left canine and P₄ of TM 292-02-01 (Fig. 1e, g) are sufficiently well preserved to be described in detail. The canine crown, which is small with an asymmetrical outline in occlusal view at the cervix level (maximum mesiodistal length is 10.0 mm, and buccolingual width is 8.5 mm), is broken apically and worn distally. The wear pattern of the lower canine indicates that occlusion of the upper canine was solely against the large distal tubercle that projects lingually. This pattern of occlusion is clearly marked by a grooved wear strip on the distal

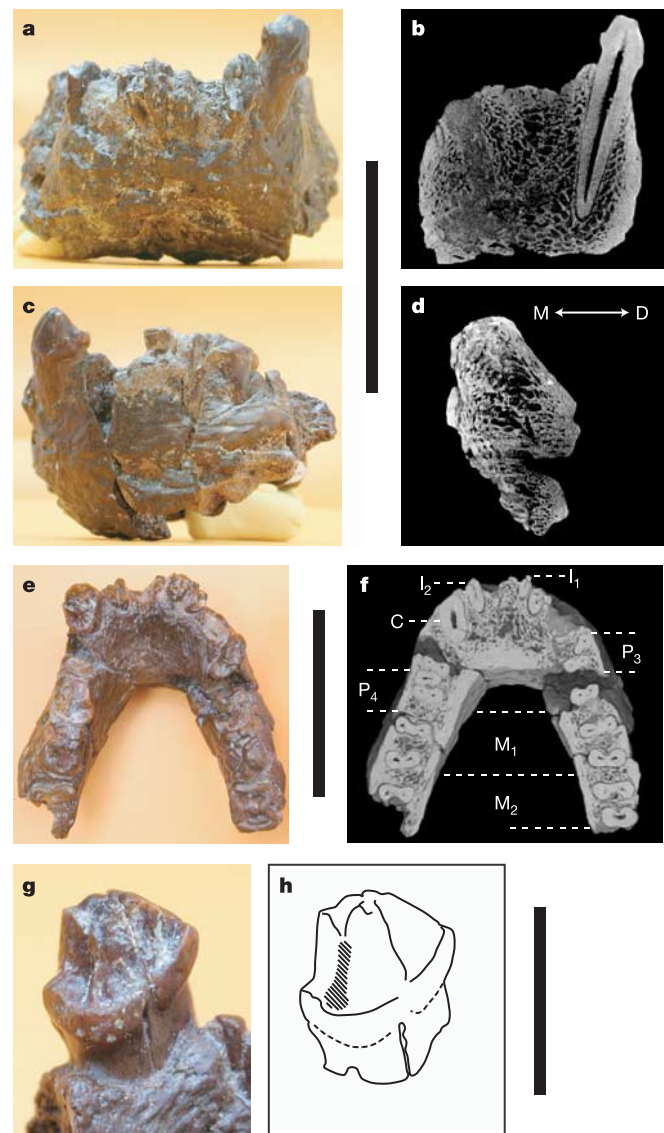


Figure 1 Lower jaw (TM 292-02-01) assigned to *Sahelanthropus tchadensis*. **a**, Frontal view. **b**, Left canine posterior coronal computed tomography (CT) scan (flipped horizontally) (scanner, University Museum, Tokyo, Japan). **c**, Left buccal view. **d**, Symphysis midsagittal CT scan (M, mesial; D, distal) (scanner, University Museum, Tokyo, Japan). **e**, Occlusal view. **f**, Three-dimensional reconstruction with axial CT scan; root pattern shown is taken just below the cervix (synchrotron, ESRF, Grenoble, France). **g**, **h**, Left canine disto-lingual view (**g**) and drawing (**h**) showing the location of the distal wear strip and indentation. Scale bar, 4 cm (**a–f**); 1 cm (**g**, **h**).

Table 1 New specimens of *Sahelanthropus tchadensis*

Specimen number	Collected	Element	Discoverer
TM 292-02-01 (Fig. 1)	2002	Mandibular fragment	MPFT
TM 247-01-02 (Fig. 2)	2001	Right mandibular corpus fragment	MPFT
TM 266-01-462 (Fig. 3)	2001	Right P ₃	MPFT

MPFT, Mission Paléoanthropologique Franco-Tchadienne.

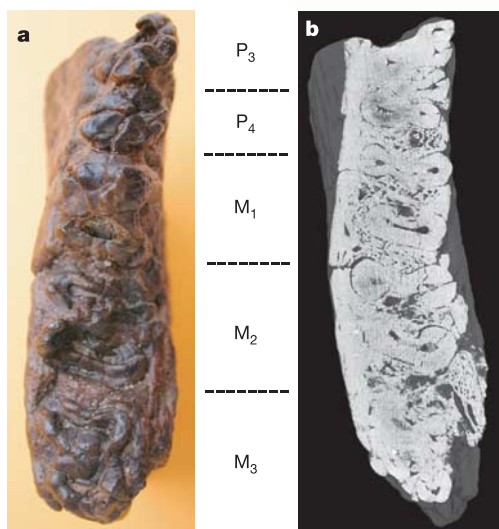


Figure 2 Right lower jaw (TM 247-01-02) assigned to *Sahelanthropus tchadensis*. **a**, Occlusal view. **b**, Three-dimensional reconstruction with axial computed tomography scan, root pattern shown is taken just below the cervix (synchrotron, ESRF, Grenoble, France). Scale bar, 4 cm.

enamel surface that terminates in an indentation on the occlusal surface of the distal shelf-like tubercle (Fig. 1g, h). TM 292-02-01 is therefore consistent with the absence of a functional C/P₃ honing complex in *S. tchadensis*³. The crown of the canine is short, yet its root is surprisingly long (Fig. 1b). Both the mesial and distal crown shoulders (Fig. 1g, h) are very low relative to the cervix. In contrast to the configuration in *Ardipithecus kadabba*^{10,11} (Fig. 1b), the mesial shoulder is only slightly more apical than the distal shoulder. A distinct marginal ridge is present on the mesiolingual surface.

The subrectangular P₄ of TM 292-02-01 has a maximum mesiodistal length of 8.0 mm, with a distolingually well-developed talonid; the elongated partly preserved trigonid has a distinct distally positioned metaconid (Fig. 1e).

TM 266-01-462 (Fig. 3) is a right P³ lacking roots and a portion of the distal intercusp crown. Dimensions of the P³ are 13.0 mm (buccolingual) and about 7.3 mm (minimum mesiodistal at paracone). The occlusal crown outline is oval with a slight concavity on its mesial surface below the marginal ridge. The mesial enamel surface shows a well-delimited interproximal canine wear facet below the mesial marginal ridge, confirming the lack of a diastema between C₁ and P₃. The mesial marginal ridge is above mid-crown level. In addition, the TM 266-01-462 premolar is bicuspid with a tall, conical paracone, and a smaller, lower protocone that is more mesially located than the paracone. Both cusps are slightly worn, with the tip of the paracone showing a small area of dentine exposure. The P³ presents a mesio-cervical enamel extension on the steeply sloping buccal surface. The small anterior fovea is mesial to the transverse crest of the paracone and bordered by a moderately thick mesial marginal ridge that slopes downwards buccally. The paracone has a prominent, rounded transverse crest extending slightly mesially to the median groove between the two cusps. The mesially facing triangular portion of the occlusal surface present in African apes and *Ardipithecus ramidus* is absent¹².

The maximum radial enamel thickness measured from micro computed tomography scans of the P³ (TM 266-01-462, protocone and paracone), upper right M² and M³ (TM 266-01-60-1, paracone, protocone and hypocone) and the right P₄ (TM 266-02-154-1, protoconid) ranges from 1.2 to 1.9 mm. The lower buccal and upper lingual cusps tend to have thicker enamel (1.4–1.9 mm) than the lower lingual and upper buccal cusps (1.2–1.6 mm). The postcanine



Figure 3 Right upper P³ (TM 266-01-462) assigned to *Sahelanthropus tchadensis*. **a**, Occlusal view. **b**, Mesial view. Scale bar, 1 cm.

cuspal enamel thickness in these *S. tchadensis* specimens is therefore intermediate between published values for chimpanzees and australopithecines¹².

The new material presented here is important for several reasons. First, the fossils add substantially to the holotype cranium, TM 266-01-60-1, which is remarkable in its completeness and preservation. The *S. tchadensis* hypodigm now includes a minimum of six individuals (a maximum of nine) from three sites in a small area of the Anthracotheriid Unit. Second, these new fossils now permit a more complete and reliable understanding of this earliest known hominid taxon. *S. tchadensis* shares major derived features with other recognized hominids that are consistent with its position in the hominid clade, close to the last common ancestor of chimpanzees and humans. In the dentition these anatomical characters are a non-honing C/P₃ complex; no diastema between C and P₃; a vertical symphysis with weak transverse tori; canines with a small crown and long root, a lower canine crown with a large distal tubercle, both shoulders being very low; an upper P³ with a steeply sloping buccal surface; postcanine teeth with maximum radial enamel thickness intermediate between chimpanzees and australopithecines; and bulbous, slightly crenulated postcanine occlusal morphology. All the hominid mandibular premolar specimens from Toros-Menalla have the same root pattern, with two roots and three separate pulp canals in each premolar (one mesial and two distal) retaining the presumed primitive condition for the *Pan/Homo* clade¹³ (Figs 1f and 2b).

The anatomical characters of the new material of *S. tchadensis*, such as a lower canine crown with a distinct mesial marginal ridge and a distal grooved wear strip ending on a large distal tubercle (a feature consistent with the absence of a honing C/P₃ complex), confirm the morphological differences of the Chadian species from African apes, and its morphological affinities with the hominid clade. Although the new fossils provide valuable data, the nearly complete cranium TM 266-01-60-1 remains a key specimen for *S. tchadensis* that is older than any other Late Upper Miocene hominid so far known^{10,11,14}. Identifiable derived features of *S. tchadensis*³ are a face with an anteroposteriorly short premaxilla, an anteriorly positioned foramen magnum linked to a short basioccipital and a sub-horizontal nuchal plane, a downward lipping of the nuchal crest, and a non-honing C/P₃ complex. Post-mortem plastic deformation of the TM 266 cranium has precluded further detailed analysis⁴. However, a virtual three-dimensional reconstruction of the TM 266 cranium (presented in ref. 15) provides additional morphological information for the more precise evaluation of its systematic position with respect to the extant great apes and to other known hominid taxa, and for testing hypotheses about key aspects of its behaviour, particularly its mode of locomotion. □

Received 17 September 2004; accepted 26 January 2005; doi:10.1038/nature03392.

1. Brunet, M. *et al.* The first Australopithecine 2,500 kilometres west of the Rift Valley (Chad). *Nature* **378**, 273–275 (1995).
2. Brunet, M. *et al.* *Australopithecus bahrelghazali*, une nouvelle espèce d'Hominidé ancien de la région de Koro Toro (Tchad). *C. R. Acad. Sci. Paris* **322**, 907–913 (1996).
3. Brunet, M. *et al.* A new hominid from the Upper Miocene of Chad, Central Africa. *Nature* **418**, 145–151 (2002).

4. Wolpoff, M., Senut, B., Pickford, M. & Hawks, J. *Sahelanthropus* or 'Sahelpithecus'? *Nature* **419**, 581–582 (2002).
5. Brunet, M. *et al.* *Sahelanthropus* or 'Sahelpithecus'? (Reply). *Nature* **419**, 582 (2002).
6. Vignaud, P. *et al.* Geology and palaeontology of the Upper Miocene Toros-Menalla hominid locality, Djurab Desert, Northern Chad. *Nature* **418**, 152–155 (2002).
7. Brunet, M. *et al.* Tchad: un nouveau site à Hominidés Pliocène. *C. R. Acad. Sci. Paris* **324**, 341–345 (1997).
8. Deino, A. L., Tauxe, L., Monaghan, M. & Hill, A. Ar⁴⁰/Ar³⁹ geochronology and paleomagnetic stratigraphy of the Lukeino and lower Chemeron Formations at Tabarin and Kapcheberek, Tugen Hills, Kenya. *J. Hum. Evol.* **42**, 117–140 (2002).
9. MacDougall, I. & Feibel, C. in *Lothagam the Dawn of Humanity in Eastern Africa* (eds Leakey, M. G. & Harris, J. M.) 43–64 (Columbia University Press, New York, 2003).
10. Haile-Selassie, Y., Suwa, G. & White, T. Late Miocene teeth from Middle Awash, Ethiopia, and early hominid dental evolution. *Science* **303**, 1503–1505 (2004).
11. Haile-Selassie, Y. Late Miocene hominids from the Middle Awash, Ethiopia. *Nature* **412**, 178–181 (2001).
12. White, T. D., Suwa, G. & Asfaw, B. *Australopithecus ramidus*, a new species of hominid from Aramis, Ethiopia. *Nature* **371**, 306–312 (1994).
13. Wood, B. A., Abbott, S. A. & Yttersthaug, H. Analysis of the dental morphology of Plio-Pleistocene hominids. IV. Mandibular postcanine root morphology. *J. Anat.* **156**, 107–139 (1988).
14. Senut, B., Pickford, M., Gommery, D., Mein, P. & Cheboi, K. First hominid from the Miocene (Lukeino Formation, Kenya). *C. R. Acad. Sci. Paris* **332**, 137–144 (2001).
15. Zollikofer, C. P. E. *et al.* Virtual cranial reconstruction of *Sahelanthropus tchadensis*. *Nature* doi: 10.1038/nature03397 (this issue).

Acknowledgements We thank the Chadian Authorities (Ministère de l'Éducation Nationale de l'Enseignement Supérieur et de la Recherche, Université de N'djaména, CNAR), the Ministère Français de l'Éducation Nationale (Faculté des Sciences, Université de Poitiers), the Ministère de la Recherche (CNRS: Département SDV & ECLIPSE), the Ministère des Affaires Étrangères (DCSUR, Paris and SCAC, N'Djaména) to the Région Poitou-Charentes, the American School of Prehistoric Research, the RHOI (co-Principal Investigators F. C. Howell and T. D. White), the Armée Française, MAM and Epervier for logistical support; the scanner staff of the University Museum, the University of Tokyo (microCT scanning, G. Suwa); to the ESRF, Grenoble (W. G. Stirling, General Director, A. Bravin and C. Nemoz, ID 17); many colleagues and friends for their help, especially G. Suwa for enamel thickness measurements, P. Tafforeau for ESRF three-dimensional scan reconstructions; T. D. White for discussions; all the other members of the Mission Paléanthropologique Franco-Tchadienne (MPFT) who joined us for field missions; S. Riffaut and X. Valentin for technical support; and G. Florent and C. Noël for administrative guidance at the MPFT.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.B. (michel.brunet@univ-poitiers.fr).

Virtual cranial reconstruction of *Sahelanthropus tchadensis*

Christoph P. E. Zollikofer¹, Marcia S. Ponce de León¹, Daniel E. Lieberman², Franck Guy^{2,3}, David Pilbeam², Andossa Likius⁴, Hassane T. Mackaye⁴, Patrick Vignaud³ & Michel Brunet³

¹Anthropologisches Institut/MultiMedia Laboratorium, Universität Zürich-Irchel, Winterthurerstrasse 190, 8057 Zürich, Switzerland

²Peabody Museum, Harvard University, 11 Divinity Avenue, Cambridge, Massachusetts 02138, USA

³Laboratoire de Géobiologie, Biochronologie et Paléontologie Humaine, CNRS UMR 6046, Faculté des Sciences, Université de Poitiers, 40 Avenue du Recteur Pineau, 86022 Poitiers Cedex, France

⁴Université de N'Djaména, BP 1117, N'Djaména, Tchad

Previous research in Chad at the Toros-Menalla 266 fossiliferous locality (about 7 million years old) uncovered a nearly complete cranium (TM 266-01-60-1), three mandibular fragments and several isolated teeth attributed to *Sahelanthropus tchadensis*^{1–3}. Of this material, the cranium is especially important for testing hypotheses about the systematics and behavioural characteristics of this species, but is partly distorted from

fracturing, displacement and plastic deformation. Here we present a detailed virtual reconstruction of the TM 266 cranium that corrects these distortions. The reconstruction confirms that *S. tchadensis* is a hominid and is not more closely related to the African great apes^{4,5}. Analysis of the basicranium further indicates that *S. tchadensis* might have been an upright biped, suggesting that bipedalism was present in the earliest known hominids, and probably arose soon after the divergence of the chimpanzee and human lineages.

Primary distortion in TM 266-01-60-1 results from morphological discontinuities along major cracks between the left and right sides of the face, between the supraorbital torus and the zygomatics, between the left and right posterior cranial vault including the nuchal plane and basioccipital, and along a coronally oriented crack between left frontal and temporoparietal portions of the vault (Fig. 1; also see Fig. 1 in ref. 1). However, anatomical continuity is well preserved in the sagittal and parasagittal planes, particularly between the face, the neurocranium and the basicranium. Anatomical continuity in the basicranium extends from the basisphenoid to the nuchal plane and within each of the cranial units delimited by major cracks, as evident from matching fracture lines between adjacent parts. Plastic deformation resulting in left–right asymmetry is noticeable in the maxilla. The fossil is barely affected by expanding matrix distortion⁶, and no missing regions need to be estimated to reconstruct its original form.

A high-resolution computed tomography scan was used to create a digital representation of the TM 266 cranium that was disassembled along major cracks, cleaned of adhering matrix with the use of digital filtering, and then reconstructed virtually with two different established protocols (see Methods). The reconstruction, illustrated in Fig. 2, was evaluated with three independent tests. First, the face and neuro-basiscranial complex, which were reconstructed separately, fitted together at multiple points in an approximately coronal plane along the superior and lateral margins of the post-orbital region. Second, the reconstructed morphology was assessed a posteriori against an anatomical constraint not considered during the virtual reconstruction. In all mammals including primates, the posterior maxillary (PM) plane is approximately perpendicular relative to the neutral horizontal axis (NHA) of the orbits⁷. PM orientation was estimated by a plane that passes, in lateral projection, from the maxillary tuberosities through the pterygopalatine fossae⁸. In the TM 266 reconstruction, this plane is about 89° relative to the NHA (estimated from the orbital margins and the partly preserved medial walls). As a third test, the TM 266 reconstruction was compared with three-dimensional shape variability in a comparative African ape/fossil hominid sample (see Methods). We performed a generalized least-squares superimposition⁹ of the symmetrized landmark configurations¹⁰ of all specimens and calculated the minimum form change necessary to transform the TM 266 reconstruction to the closest possible hypothetical *Pan* and *Gorilla* cranial forms with the use of the 99% probability density borders as a minimum-distance criterion (Fig. 3). Figure 3a–c shows this procedure for the first three PCs, which account for more than 58% of the total shape variability. To account for allometric shape effects, all shape PCs were regressed against centroid size to obtain a common allometric shape score¹¹ (Fig. 3d). The isolated fragments of the TM 266 cranium were then positioned to fit the calculated three-dimensional landmark configurations of the closest-possible *Pan* and *Gorilla* shapes (Fig. 3e). The resulting 'Pan-like' and 'Gorilla-like' morphologies are anatomically infeasible, involving overlap between neurocranial fragments and disruption of anatomical continuity between neighbouring facial fragments. Although the cranial morphology of TM 266-01-60-1 cannot be reconstructed to fall within the size–shape

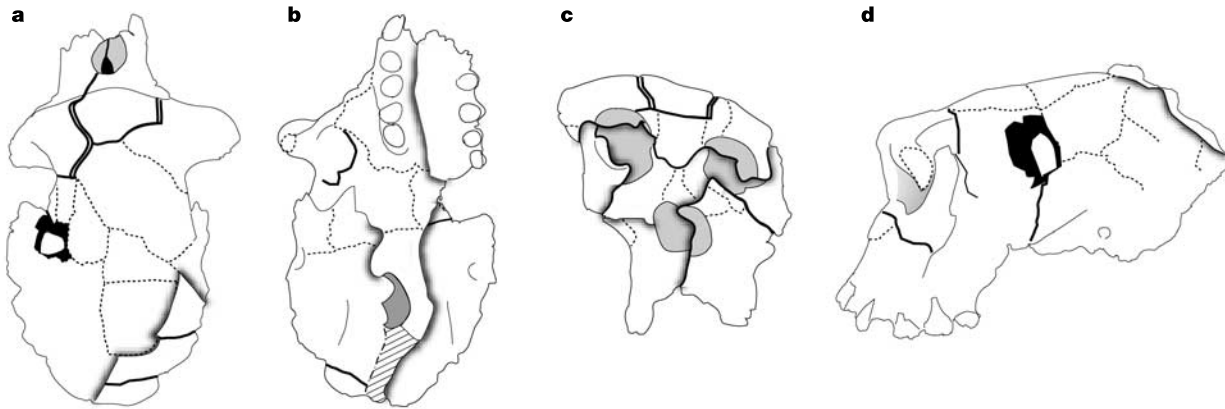


Figure 1 State of preservation of the unreconstructed TM 266 cranium. **a**, Superior view. **b**, Inferior view. **c**, Frontal view. **d**, Left lateral view. Thin lines and grey areas indicate anatomical structures; shadowed lines, superposition of cranial parts resulting from postmortem distortion; bold lines, major cracks (double lines, disruption of anatomical

continuity); stippled lines, matching fracture lines between electronically isolated components; black areas, missing parts or matrix filling. Note that the everted occipital fragment (hatched area in **b**) obscures the original morphology and orientation of the nuchal plane.

space of known African ape morphologies, it is within the size–shape space defined by other Pliocene hominids.

The reconstruction in Fig. 2 is therefore a robust estimate of the cranial form of TM 266-01-60-1 that supports most of the details originally described^{1,2}. However, several features differ notably from those of the original specimen: the cranium as a whole is wider, the occipital contour is rounder sagittally, the nuchal plane is oriented more horizontally, the orbits are larger and more circular, and the face is superoinferiorly taller (additional standard craniometric measurements are provided in Supplementary Table 1). The changes evident in the TM 266 reconstruction highlight its unique morphology and confirm several derived features shared with later

hominids such as a relatively vertical face with an anteroposteriorly short premaxilla; an anteriorly-positioned foramen magnum linked to a relatively short basioccipital; a relatively flat, large, and horizontally-oriented nuchal plane; and downward lipping of the nuchal crest^{12–14}. These features, together with other dental features (see refs 1, 2), support the conclusion that *Sahelanthropus* is a hominid (*contra* Wolpoff *et al.*^{4,5}).

Finally, the TM 266 reconstruction permits an assessment of the hypothesis that *Sahelanthropus* was a biped, an important feature of Pliocene hominids and possibly several Late Miocene hominids^{15,16}. Unequivocal evidence for bipedalism is difficult to obtain from the cranium, but several lines of evidence suggest that TM 266-01-60-1

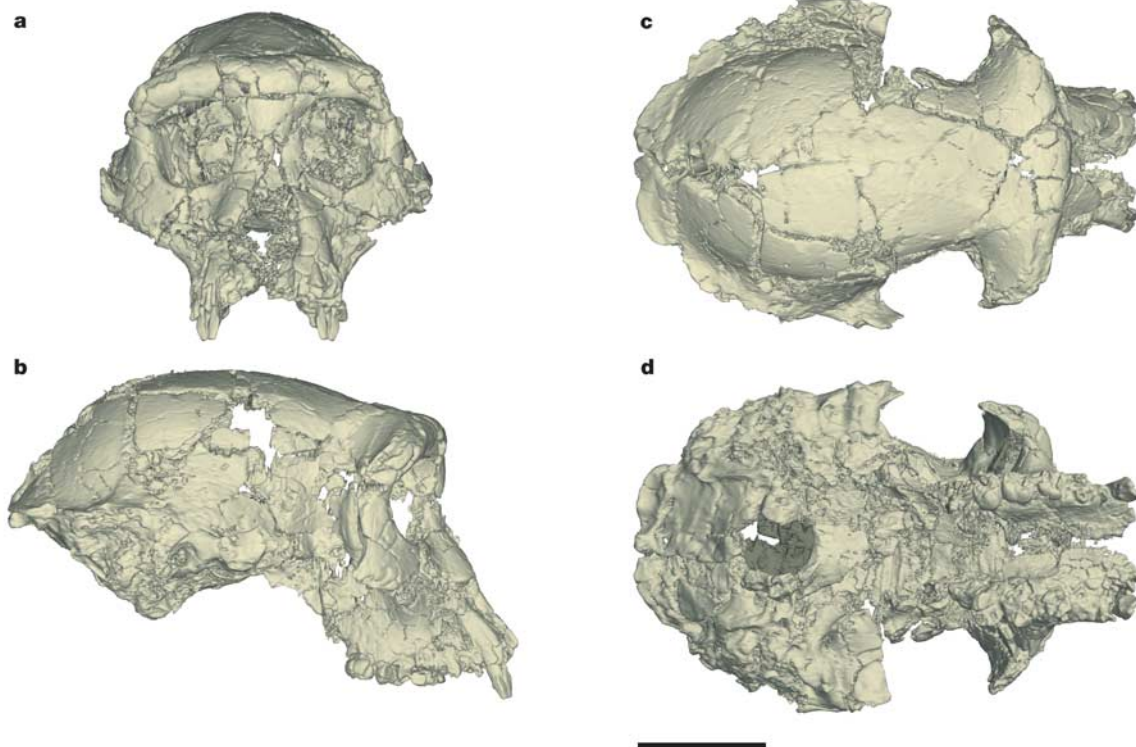


Figure 2 Virtual reconstruction of the TM 266 cranium (Frankfurt Horizontal plane orientation, orthographic projection). **a**, Frontal view. **b**, Right lateral view. **c**, Superior view. **d**, Inferior view. Scale bar, 5 cm.

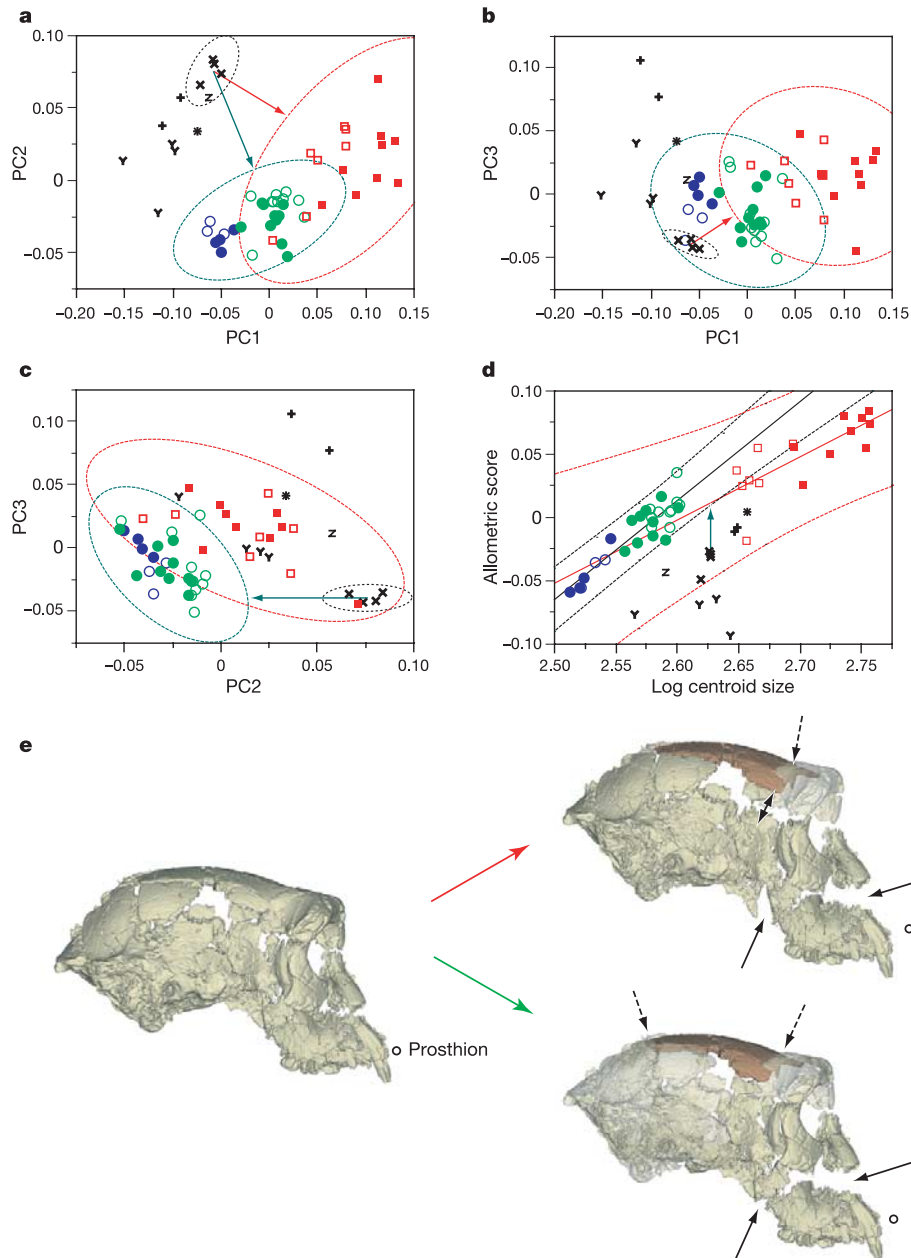


Figure 3 TM 266 cranial reconstruction and comparative fossil hominid/African ape morphology. **a–d**, Evaluation of size-adjusted minimum distance between TM 266-01-60-1 and the African apes with data scatter in size–shape space defined by PCs1–3. Variance proportions: PC1, 39.1%; PC2, 11.3%; PC3, 8.3%. X, reconstructed variants of TM 266-01-60-1; Y, fossil *Homo*; Z, *Australopithecus africanus*; stars, *A. afarensis*; plus signs, *Paranthropus*; blue circles, *Pan paniscus*; green circles,

Pan troglodytes; squares, *Gorilla gorilla* (open symbols, females; filled symbols, males). Stippled lines indicate 99% probability densities for *Pan*, *Gorilla* and TM 266-01-60-1 reconstructions. **e**, Transformation of TM 266-01-60-1 into hypothetical closest African ape morphologies in multidimensional size–shape space: green arrows, transformation into *Pan*; red arrows, transformation into *Gorilla*; black arrows indicate resulting disruption (solid) and overlap (dashed) between neighbouring fragments.

might have been bipedal. Despite substantial differences in neck orientation, humans and non-human primates tend to locomote with their orbital planes (the line joining the superior and inferior margins of the orbits) approximately perpendicular to the ground¹⁷. In addition, primates orient the upper cervical vertebrae approximately perpendicular to the plane of the foramen magnum, and with only a limited range (about 10°) of flexion and extension possible at the cranio-cervical joint¹⁸. The combined effect of these angular constraints is that the angle between the foramen magnum and the orbital plane (Fig. 4) is nearly perpendicular in *Homo sapiens* ($103.2 \pm 6.9^\circ$,

$n = 23$) but more acutely angled in *Pan troglodytes* ($63.7 \pm 6.2^\circ$, $n = 20$), and other species with more pronograde postures. The foramen magnum angle relative to the orbital plane in the TM 266 reconstruction is 95° , similar to that in humans and later bipedal hominids such as *Australopithecus afarensis* (AL 444-2) and *A. africanus* (Sts 5)^{13,17}. TM 266-01-60-1 as a quadruped would require an unusually extended angle of the neck relative to the plane of the foramen magnum.

Although increases in brain volume relative to cranial base length have been implicated in horizontal rotation of the posterior cranial base¹⁹, such an explanation is unlikely for the TM 266 cranium

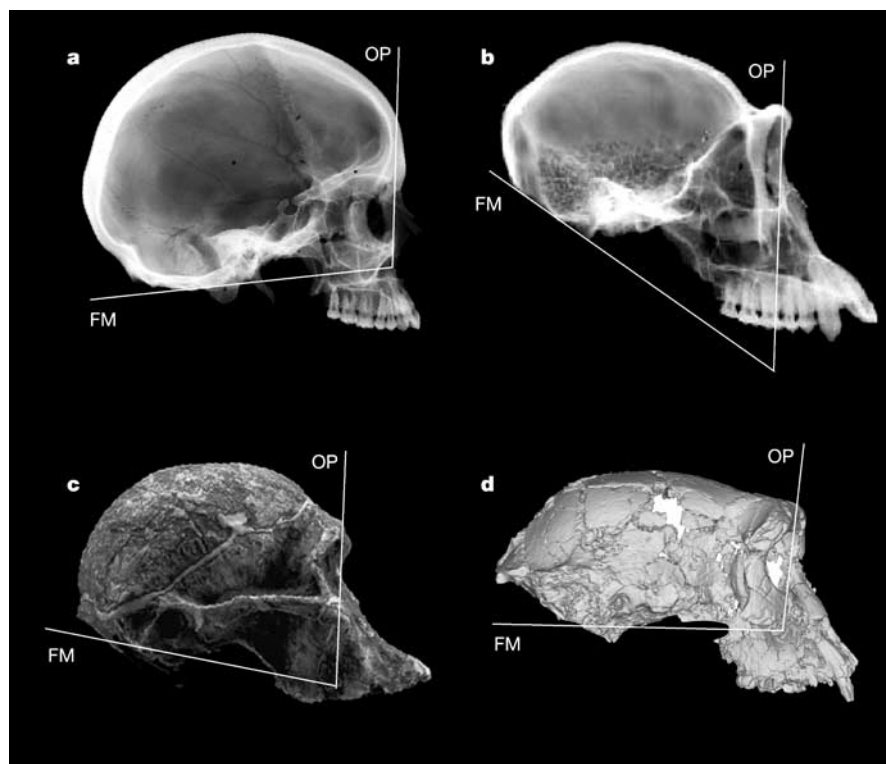


Figure 4 Angular relationship between foramen magnum (FM) and orbital (OP) planes. **a**, *Homo sapiens*; **b**, *Pan troglodytes*; **c**, *Australopithecus africanus* (Sts 5); **d**, TM 266 reconstruction.

whose estimated endocranial volume, 360–370 ml, is the smallest yet documented for an adult hominid but is within the range in chimpanzees²⁰. Another related indication of bipedality in *S. tchadensis* is the flat nuchal plane oriented at about 36° relative to the Frankfurt Horizontal, well within the range of *Australopithecus* and *Homo* but not *Pan*²¹. In addition, the nuchal crest of the TM 266 cranium has downward lipping, a feature present in other bipedal hominids (for example AL 444-2, OH 5) but not in *Pan* or *Gorilla*. However, postcranial evidence will be necessary to test more rigorously the hypothesis that *S. tchadensis*—the earliest known hominid, found 2,600 km west of the East African rift valley—was a biped. □

Methods

Data acquisition and state of preservation

The TM 266-01-60-1 original was scanned with an Industrial Systems CT scanner (tube voltage 450 kV, tube current 5 mA, beam collimation 0.4 mm, interslice distance 0.4 mm, pixel size 0.2 mm × 0.2 mm, pixel depth 16 bits). The virtual cranium was disassembled along major cracks, and matrix filling was removed from endocranial and paranasal cavities with the use of interactive data segmentation tools²². Virtual separation of cranial fragments revealed substantial overlap between right and left sides of the cranium from post-mortem compression, causing an elongated appearance of the vault in superior view (see figure 1c in ref. 1). In addition, the lower face and orbital margins are shifted left and superiorly relative to the supraorbital torus and zygomatic processes of the frontal (Fig. 1c), the basioccipital is shifted towards the left petrosal (Fig. 1b), and the right posterior cranial vault and nuchal plane overlay the left side (Fig. 1a, b), altering the true orientation of the nuchal plane in the sagittal plane (see figure 1b in ref. 1). The exposed borders of parts of the cranial vault and palate are partly eroded; the mastoids and petrosal portions of the temporal bones also suffered some surface damage but are undistorted, as evinced by the mirror-symmetric position of the well-preserved left and right inner ear cavities relative to features such as the external acoustic meatus and the stylomastoid foramen. The left temporal fossa is partly missing, but corresponding structures on the right side are well preserved.

Virtual three-dimensional reconstruction

The reconstruction of the cranium followed established methods^{23,24}. Once partitioned, isolated fragments were repositioned and reoriented in virtual space to restore morphological continuity along fractures, sutures and other anatomical features within

and between bones. The cranium was independently reconstructed four times by two of us (M.S.P.L. and C.P.E.Z.), each using two different protocols. Protocol A used features shared by all mammal crania to position and orient each fragment. First, the basioccipital was positioned and oriented in the midsagittal plane. The temporals were then adjoined from both sides and aligned by placing all of the left and right semicircular canals in approximately parallel orientation^{25,26}. Lateral and superior parts of the vault were adjoined by using the well-preserved temporal lines to establish bilateral symmetry. Within the face, displaced but undistorted portions of the supraorbital torus and orbital margins were repositioned symmetrically relative to the midsagittal plane. Left–right asymmetry in the maxilla from plastic taphonomic deformation was partly corrected with the use of published methods²⁴. Protocol B used a geometric approach based on stepwise reduction of degrees of freedom of the position and orientation of individual parts relative to each other. This method takes advantage of the almost complete preservation of the TM 266 cranium, in which the position and orientation of each fragment is spatially constrained by contacts with all neighbouring fragments, and overall morphology is constrained by bilateral symmetry. Translational degrees of freedom were first reduced by re-establishing morphological continuity between dislocated fragments along matching fracture lines (along the nuchal plane, along cracks in the right parietal, between parts of the supraorbital torus, and between dislocated parts of the midface). Rotational degrees of freedom between adjacent fragments were then reduced by stepwise integration of fragments into the reconstruction, followed by iterative adjustments until a symmetrical integrated morphology was achieved. These procedures were applied to orient left and right neurobasiscranial sides relative to each other, and the maxillae relative to the midface. Last, deviations from bilateral symmetry in the maxilla were partly corrected as in protocol A.

In both protocols, the face and the braincase were reconstructed independently and then assembled by using anatomical continuities within the squamous portions of the frontal; along preserved continuities between the basisphenoid, the pterygoid processes and the right side maxillary tuberosity; and between the bones of the right temporal fossa (squamous sphenoid, zygomatic, maxilla and frontal). Differences between the four reconstructions, as visualized in shape space, are comparatively small (Fig. 3a–c) and reflect inter-observer and inter-protocol disparity. Major variations concern maxillary width measured at M² (±0.9 mm); foramen magnum height relative to porion (±1.1 mm); and facial orientation relative to the braincase, as measured by the angle between nasion–basion and nasion–prosthion (±2.3°). Figure 2 shows the final result, obtained by averaging all four reconstructions.

Geometric morphometric analysis

The analysis included adult crania of 16 *G. gorilla* (9 males, 7 females), 20 *P. troglodytes* (10 males, 10 females), 7 *P. paniscus* (4 males, 3 females) and 8 fossil hominids (casts of ER1813, ER1470, ER3733, KNM-WT 15000, STS5, OH5, ER406 and virtual reconstruction of AL444-2). All primate specimens are from the A. H. Schultz Collection,

University of Zurich, the Peabody Museum, Harvard University, and the Royal Africa Museum, Tervuren. Midsagittal landmarks used were nasion, glabella, bregma, lambda, inion, opisthion, basion, sphenobasion, staphylion, prosthion and nasospinale; paired landmarks (left and right sides taken when possible) were maxillofrontale, supraorbitale, orbitale, frontomale orbitale, zygomaxillare, jugale, foramen infraorbitale, M² most buccal point, foramen stylomastoideum, foramen caroticum, foramen ovale, asterion, porion and pterion.

Received 17 September 2004; accepted 25 January 2005; doi:10.1038/nature03397.

1. Brunet, M. *et al.* A new hominid from the Upper Miocene of Chad, Central Africa. *Nature* **418**, 145–151 (2002).
2. Brunet, M. *et al.* New material of the earliest hominid from the Upper Miocene of Chad. *Nature* doi:10.1038/nature03392 (this issue).
3. Vignaud, P. *et al.* Geology and palaeontology of the Upper Miocene Toros-Menalla hominid locality, Chad. *Nature* **418**, 152–155 (2002).
4. Wolpoff, M. H., Senut, B., Pickford, M. & Hawks, J. *Sahelanthropus* or '*Sahelpithecus*'? *Nature* **419**, 581–582 (2002).
5. Brunet, M. *et al.* *Sahelanthropus* or '*Sahelpithecus*'? (Reply). *Nature* **419**, 582 (2002).
6. White, T. Early hominids—diversity or distortion? *Science* **299**, 1994–1997 (2003).
7. McCarthy, R. C. & Lieberman, D. E. Posterior maxillary (PM) plane and anterior cranial architecture in primates. *Anat. Rec.* **264**, 247–260 (2001).
8. Enlow, D. H. *Facial Growth* (Saunders, Philadelphia, 1990).
9. Rohlf, F. J. & Slice, D. Extensions of the Procrustes method for the optimal superimposition of landmarks. *Syst. Zool.* **39**, 40–59 (1990).
10. Zollikofer, C. P. E. & Ponce de León, M. S. Visualizing patterns of craniofacial shape variation in *Homo sapiens*. *Proc. R. Soc. Lond. B* **269**, 801–807 (2002).
11. Penin, X., Berge, C. & Baylac, M. Ontogenetic study of the skull in modern humans and the common chimpanzees: neotenic hypothesis reconsidered with a tridimensional Procrustes analysis. *Am. J. Phys. Anthropol.* **118**, 50–62 (2002).
12. Aiello, L. & Dean, C. *An Introduction to Human Evolutionary Anatomy* (Academic, London, 1990).
13. Kimbel, W. H., Johanson, D. & Rak, Y. *The Skull of Australopithecus afarensis* (Oxford Univ. Press, New York, 2004).
14. Wood, B. & Richmond, B. G. Human evolution: taxonomy and paleobiology. *J. Anat.* **196**, 19–60 (2000).
15. Haile-Selassie, Y. Late Miocene hominids from the Middle Awash, Ethiopia. *Nature* **412**, 178–181 (2001).
16. Pickford, M., Senut, B., Gommery, D. & Treil, J. Bipedalism in *Orrorin tugenensis* revealed by its femora. *C. R. Palevol.* **1**, 191–203 (2002).
17. Strait, D. S. & Ross, C. F. Kinematic data on primate head and neck posture: implications for the evolution of basicranial flexion and an evaluation of registration planes used in paleoanthropology. *Am. J. Phys. Anthropol.* **108**, 205–222 (1999).
18. Graf, W., de Waele, C. & Vidal, P. P. Functional anatomy of the head–neck movement system of quadrupedal and bipedal mammals. *J. Anat.* **186**, 55–74 (1995).
19. Lieberman, D. E., Ross, C. F. & Ravosa, M. J. The primate cranial base: ontogeny, function, and integration. *Yb. Phys. Anthropol.* **63**, 117–169 (2000).
20. Holloway, R. L., Broadfield, D. C., Yuan, M. S., Schwartz, J. H. & Tattersall, I. *The Human Fossil Record Vol. 4: Brain Endocasts* (Wiley, New York, 2004).
21. Kimbel, W. H., White, T. D. & Johanson, D. C. Cranial morphology of *Australopithecus afarensis*: a comparative study based on a composite reconstruction of the adult skull. *Am. J. Phys. Anthropol.* **64**, 337–388 (1984).
22. Gonzalez, R. C. & Woods, R. E. *Digital Image Processing* (Prentice Hall, Upper Saddle River, New Jersey, 2002).
23. Zollikofer, C. P. E., Ponce de León, M. S. & Martin, R. D. Computer-assisted paleoanthropology. *Evol. Anthropol.* **6**, 41–54 (1998).
24. Ponce de León, M. S. & Zollikofer, C. P. E. New evidence from Le Moustier. 1: Computer-assisted reconstruction and morphometry of the skull. *Anat. Rec.* **254**, 474–489 (1999).
25. Delattre, A. & Fenart, R. *L'hominisation du Crâne Étudiée par la Méthode Vestibulaire* (CNRS, Paris, 1960).
26. Zollikofer, C. P. E., Ponce de León, M. S., Martin, R. D. & Stucki, P. Neanderthal computer skulls. *Nature* **375**, 283–285 (1995).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the Chadian Authorities (Ministère de l'Éducation Nationale de l'Enseignement Supérieur et de la Recherche, Université de N'Djaména, Centre National d'Appui à la Recherche au Tchad), the Ministère Français de l'Éducation Nationale (Faculté des Sciences, Université de Poitiers), Ministère de la Recherche (CNRS: Département SDV & ECLIPSE), Ministère des Affaires Étrangères (DCSUR Commission des fouilles, Paris, et SCAC Ambassade de France à N'Djaména), the Revealing Hominid Origins Initiative (co-principal investigators F. C. Howell and T. D. White), the American School of Prehistoric Research, Harvard University, and the Swiss National Science Foundation, for support; the MultiMedia Laboratorium, University of Zürich (P. Stucki); EMPA Dübendorf, Switzerland, for industrial computed tomography (A. Flisch); all the MPFT members; and S. Riffaut, X. Valentin, G. Florent and C. Noël for technical support and administrative guidance.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.B. (michel.brunet@univ-poitiers.fr).

Evolutionary diversification of TTX-resistant sodium channels in a predator–prey interaction

Shana L. Geffeney¹, Esther Fujimoto^{1*}, Edmund D. Brodie III², Edmund D. Brodie Jr¹ & Peter C. Ruben¹

¹Department of Biology, Utah State University, Logan, Utah 84322-5305, USA

²Department of Biology, Indiana University, Bloomington, Indiana 47405-3700, USA

* Present address: Department of Neurobiology and Anatomy, University of Utah School of Medicine, Salt Lake City, Utah 84123-3401, USA

Understanding the molecular genetic basis of adaptations provides incomparable insight into the genetic mechanisms by which evolutionary diversification takes place. Whether the evolution of common traits in different lineages proceeds by similar or unique mutations, and the degree to which phenotypic evolution is controlled by changes in gene regulation as opposed to gene function, are fundamental questions in evolutionary biology that require such an understanding of genetic mechanisms^{1–3}. Here we identify novel changes in the molecular structure of a sodium channel expressed in snake skeletal muscle, tsNa_v1.4, that are responsible for differences in tetrodotoxin (TTX) resistance among garter snake populations coevolving with toxic newts⁴. By the functional expression of tsNa_v1.4, we show how differences in the amino-acid sequence of the channel affect TTX binding and impart different levels of resistance in four snake populations. These results indicate that the evolution of a physiological trait has occurred through a series of unique functional changes in a gene that is otherwise highly conserved among vertebrates.

Identifying the connection between genotype and phenotype is the critical step uniting evolutionary studies of phenotypic diversity with more proximate investigations of development, function and structure^{5–8}. One emerging picture is that the process of adaptive radiation might be more predictable at the genetic level than previously imagined and that the regulation of genes of major effect rather than the alteration of functional products might explain much phenotypic diversity^{1,3}. Identifying the genetic basis of bill differences in Galapagos finches⁶ and eyespot patterns in butterflies⁵, and the loss of pelvic spines in sticklebacks⁷ or eyes in cave-dwelling fish⁸, has led to an increased understanding about how evolution generates convergent and divergent traits at the molecular level. In each of these cases, ecologically important morphological differences between lineages are explainable through changes in the regulation of single genes in developing tissues. However, functional changes in the protein-coding regions of structural genes explain adaptive differences in other cases, including cryptic pigmentation in rodents⁹ and pesticide resistance in insects¹⁰.

Coevolution between the garter snake *Thamnophis sirtalis* and its toxic prey, the newt *Taricha granulosa*, has resulted in geographic variability in a physiological trait, resistance to TTX in predator lineages⁴. Tetrodotoxin causes paralysis and death by binding to the outer pore of voltage-gated sodium channels and blocking nerve and muscle fibre activity¹¹. Some populations of *Taricha* have extremely high levels of TTX in their skin that provide an almost impenetrable defence against predators¹², yet elevated TTX resistance has evolved at least twice within the radiation of *T. sirtalis*^{4,13}. This physiological adaptation is at least partly accounted for by the expression of TTX-resistant sodium channels in the skeletal muscle of resistant garter snakes¹⁴. Here we identify the molecular mechanisms underlying this diversification.

The extent to which TTX affects nerve and muscle tissue is

determined by the sodium channel isoforms expressed in the tissue. TTX-sensitive members of the gene family that encodes voltage-gated sodium channels are distinguished by the presence of an aromatic amino acid in the outer pore of domain I that is responsible for high-affinity TTX binding¹⁵. In tetrodotoxic animals, TTX resistance has been linked to the substitution of non-aromatic amino acids at this critical position in domain I in TTX-sensitive members of the sodium channel gene family^{16,17}, although mutational analysis has shown that substitutions at other positions in the outer pore also can alter TTX binding^{18,19}. In rats, the TTX sensitivity of skeletal muscle tissue is altered during development and after denervation as a result of changes in the expression levels of TTX-resistant (rNav1.5) and TTX-sensitive (rNav1.4) members of the sodium channel gene family^{20,21}. TTX resistance in garter snakes might be due to a change in sodium channel gene expression²² or mutations in TTX-sensitive members of the sodium channel gene family²³.

We identified the TTX-resistant sodium channel gene expressed in skeletal muscle tissue by making a complementary DNA library from skeletal muscle of a snake from a population with elevated TTX resistance ('Benton'). We screened the library with probes designed to identify clones with skeletal and cardiac muscle sodium channel sequences (Nav1.4 and Nav1.5) and sequenced the positive cDNA library clones revealing one open-reading frame of 5,625 nucleotides that encodes an 1,875-amino-acid protein. Phylogenetic analysis of the protein-coding region of the nucleotide sequence revealed a well-supported association of our sequence with other Nav1.4 genes (see Supplementary Fig. S1 and Supplementary Methods) and we refer to this channel as tsNav1.4. The predicted protein sequence of tsNav1.4 has a tyrosine residue in the domain I position that is responsible for high-affinity TTX binding¹⁵. However, two novel amino-acid substitutions were found in other regions important for TTX binding and pore structure, the pore helix and β -strand of the domain IV outer pore^{19,24} (Fig. 1).

Sequences of tsNav1.4 from three other snake populations (Fig. 1a) with different levels of resistance to TTX all have tyrosine in the domain I position responsible for high-affinity TTX binding.

There are no sequence differences between the four populations in regions encoding the pore helix, selectivity filter and β -strand of domains I, II and III. The predicted amino-acid sequence of the pore helix, selectivity filter and β -strand of domain IV from a non-resistant population (Bear Lake) matches that of human and rat Nav1.4, indicating that it might be a TTX-sensitive channel. However, sequences from the resistant snake populations have unique amino-acid substitutions in the domain IV outer pore (Fig. 1). Because we sequenced cDNA from mRNA transcripts, we cannot eliminate the possibility that these differences reflect post-transcriptional modifications. However, alternative splicing has never been detected for Nav1.4 (refs 25, 26) and is unlikely to explain the pattern of substitutions observed in different populations.

To test whether observed differences in the amino-acid sequence alter the TTX-sensitivity of tsNav1.4, we expressed the channels in *Xenopus* oocytes. We tested both tsNav1.4 and human-snake chimaeras in which the human Nav1.4 sequence had the regions that form the outer pore, the S5–S6 linkers from domains I–IV, replaced with tsNav1.4 sequence (see Supplementary Fig. S2). We began with constructs with the tsNav1.4 sequence from Benton, then mutated individual amino acids to match pore substitutions in other snake populations, and measured the effects of these changes in the domain IV pore sequence on TTX sensitivity.

Channels with snake sequences were not more resistant to TTX block than chimaeric channels, indicating that regions outside the S5–S6 linkers might not increase the TTX resistance of snake channels (Fig. 2a, Table 1). Chimaeric channels with different domain IV pore sequences had markedly different TTX-binding affinities (Fig. 2b, d, Table 1). The TTX concentration at which 50% of sodium channels are blocked by TTX (K_d) for a chimaeric channel with the Bear Lake sequence was comparable to K_d values for TTX-sensitive Nav1.4 from rat (5 nM TTX) and human (25 nM TTX)²⁷. The K_d for the Willow Creek chimaera was three orders of magnitude greater than the K_d for Bear Lake and greater than the K_d values for TTX-resistant, mammalian cardiac channels, Nav1.5 (rat, 2.0 μ M TTX; human, 6.0 μ M TTX)²⁷. The binding affinities for the

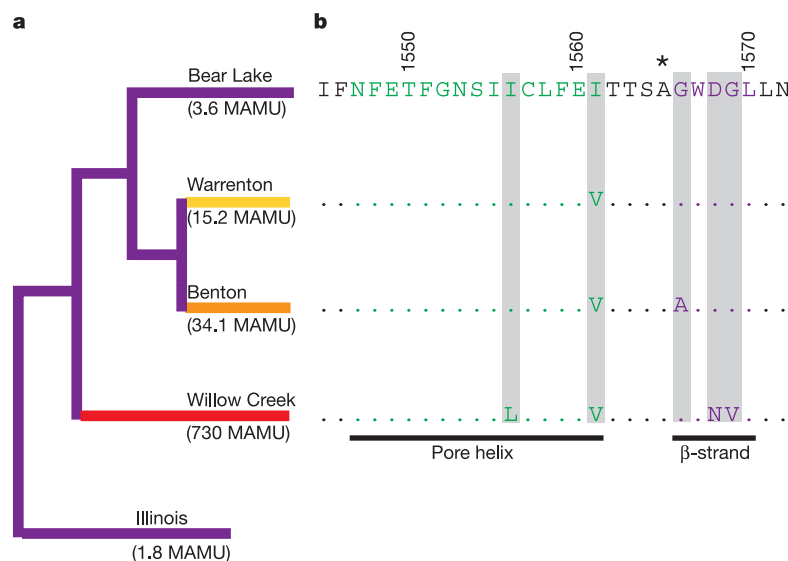


Figure 1 Amino-acid sequence differences for four snake populations. **a**, Phylogeographic relationships based on mitochondrial DNA analysis of 19 North American populations of *Thamnophis sirtalis* indicate separate origins of elevated resistance to TTX in the Willow Creek population compared with populations from Benton and Warrenton^{4,13}. Bear Lake is from a third lineage and is not resistant to TTX. Whole-animal TTX resistance for each population is reported in mass-adjusted mouse units (MAMU); branch colours reflect statistically distinguishable levels of resistance⁴.

Whole-animal TTX resistance was measured as a mass-adjusted dose of TTX (MAMU) that produced an average of 50% decrease in snake sprint speed in each population. **b**, Amino-acid alignment of part of the domain IV S5–S6 linker that affects TTX binding from tsNav1.4. Green, pore α -helix; purple, β -strand; asterisk, selectivity filter. Dots indicate identical amino acids and grey shading highlights sequence differences between populations. Despite independent evolutionary histories, all resistant snakes share the substitution of valine for isoleucine at position 1,561.

Warrenton and Benton chimaeras were intermediate to those of the other populations (Fig. 2b, Table 1). The TTX binding affinities of cloned channels with the amino-acid sequences found in the four populations are closely correlated with skeletal-muscle-fibre and whole-animal resistance of snakes from the same populations (compare Fig. 1a with Fig. 3)¹⁴. This correlation indicates that functional changes in tsNav_v1.4 are largely responsible for the differences in TTX resistance observed in snake populations.

The functional consequences of the amino-acid substitutions can be understood within the context of recent models of pore structure. A hydroxyl group on TTX interacts with an aspartic acid residue in the domain IV outer pore of rat Nav_v1.4 (ref. 28). When this charged amino acid is neutralized by substituting asparagine at this position, the TTX binding affinity of the channel is decreased by two orders of magnitude^{28,29}. Charge neutralization of the equivalent residue (position 1,568) in the Willow Creek sequence, in which aspartic acid is changed to asparagine, might similarly affect the electrostatic interaction between TTX and this pore residue²⁸. The amino-acid changes in Warrenton and Benton might affect TTX binding by altering pore structure. The substitutions are in a region of the S5–S6 linker where the protein forms a sharp bend and interactions between hydrophobic residues from the pore helix and the β -strand stabilize the structure of the mouth of the pore²⁴. Even a single change from isoleucine to valine at position 1,561 in the domain IV pore helix doubles the K_d of the Warrenton chimaera compared with that of the Bear Lake chimaera (Fig. 2b, Table 1). When this

Table 1 K_d values and 95% confidence limits for all channels

Channel type	$K_d \pm 95\%$ confidence limits (nM)
Bear Lake tsNav _v 1.4	20 ± 5.8
Bear Lake chimaera	22 ± 6.6
Benton tsNav _v 1.4	120 ± 41
Benton chimaera	210 ± 60
Warrenton chimaera	44 ± 4.2
Willow Creek chimaera	$12,000^{+2,000}_{-2,900}$
Willow Creek chimaera V1561I	$5,900 \pm 1,400$

The TTX concentration that blocked 50% of the channels (K_d) for each channel type was calculated from pooled channel data (see Methods).

valine at position 1,561 is restored to isoleucine in the Willow Creek chimaera, the K_d value of the channel is halved (Fig. 2c, Table 1).

Our results suggest that elevated TTX resistance has repeatedly evolved in snake populations through a variety of mutations in the gene that encodes tsNav_v1.4 and that these mutations alter the TTX-binding affinity of the channel by several orders of magnitude. Some of the changes observed here are unique to specific lineages (for example, glycine to alanine at position 1,566 in the Benton population, and three of the four substitutions observed in the Willow Creek population) and confer very large differences in TTX binding affinity. One mutation, which alters isoleucine to valine at position 1,561, might represent parallel evolution at that site in resistant populations, although the functional effect of the substi-

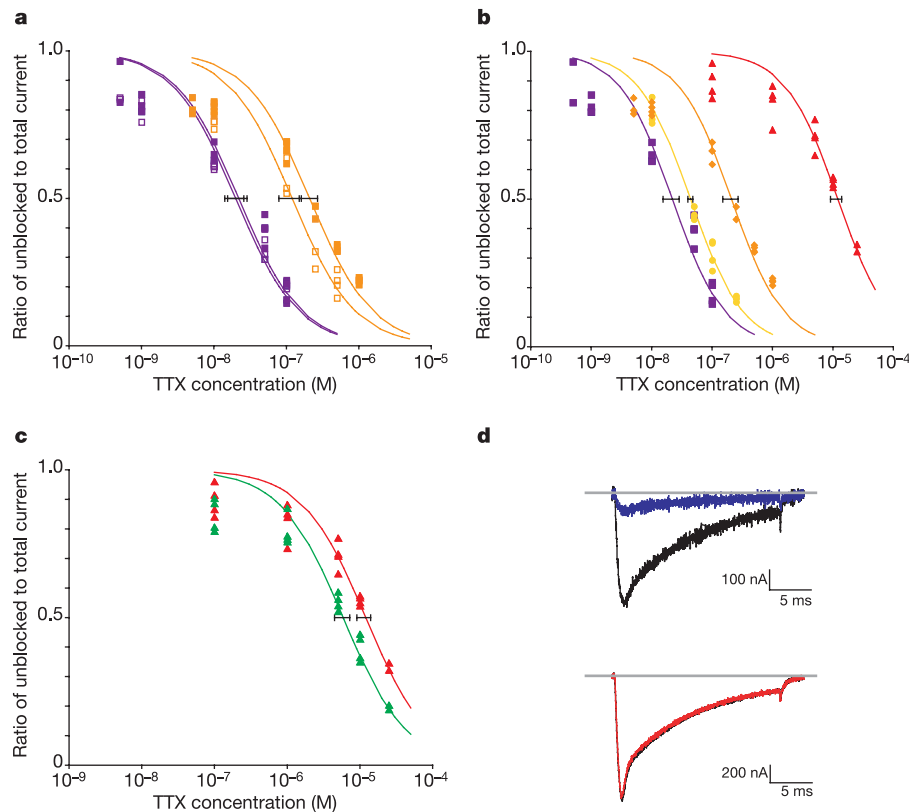


Figure 2 The effect of different TTX concentrations on tsNav_v1.4 and snake–human chimaeric channels. **a–c**, Each symbol corresponds to the ratio of unblocked to total current for an oocyte expressing the indicated channel and exposed to one TTX concentration. The TTX concentration that blocked 50% of the channels (K_d) for each channel type was calculated from pooled channel data (see Methods). K_d values ($\pm 95\%$ confidence limits) are shown for each channel type with a black bar. The lines represent the equation fitted to the data with the estimated K_d for each channel type. **a**, Channels with entirely snake sequence (open squares) are not more resistant to TTX block than chimaeric channels (filled squares). Purple, Bear Lake; orange, Benton. **b**, Chimaeric

channels with the domain IV S5–S6 linker sequence of each of the four populations are blocked by different TTX concentrations. Purple, Bear Lake; yellow, Warrenton; orange, Benton; red, Willow Creek. **c**, Changing a valine (red points and line) to an isoleucine (green points and line) at position 1,561 in the domain IV S5–S6 linker of the Willow Creek chimaera halves the K_d value of the channel. **d**, Current recordings from chimaeric channels expressed in *Xenopus* oocytes before (black lines) and after (red and blue lines) the addition of 100 nM TTX to the external solution. Channels with the domain IV S5–S6 linker sequences of Bear Lake (top panel, blue) are blocked by 100 nM TTX and Willow Creek chimaeric channels (bottom panel, red) are not. Grey bar, zero current.

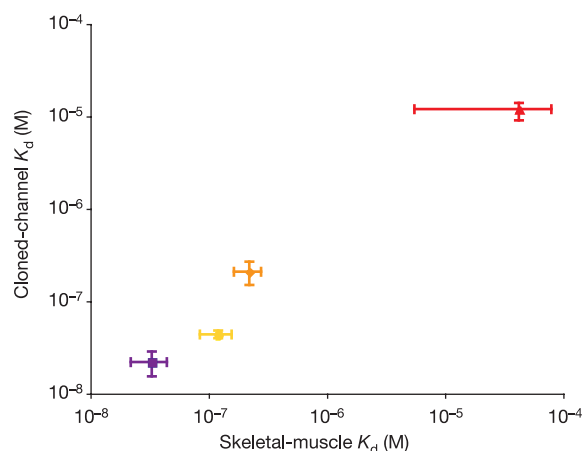


Figure 3 TTX binding affinity of cloned channels compared with skeletal muscle in four snake populations. TTX binding affinity is measured as K_d , the TTX concentration required to block 50% of channels. K_d values ($\pm 95\%$ confidence limits) for skeletal muscle were estimated from action potential rise rates recorded in skeletal muscle fibres of snakes from the four populations¹⁴. K_d values ($\pm 95\%$ confidence limits) for cloned channels were estimated from currents recorded from human–snake chimaeric channels expressed in *Xenopus* oocytes with the Bear Lake (purple), Warrenton (yellow), Benton (orange) and Willow Creek (red) sequences in the domain IV S5–S6 linker. Linear regression indicates a significant relationship between the K_d values measured from the cloned channels and skeletal muscle fibres ($F(1,3) = 17,197.116$, $R^2 = 1$, $P < 0.0001$).

tution is too small (only a twofold change in binding affinity) to explain much of the phenotypic differences observed. Our current data do not allow us to determine unequivocally whether the valine is derived or ancestral at this site (Fig. 1), although the isoleucine at this position is conserved in all other known vertebrate $\text{Na}_v1.4$ sequences. Additionally, the ancestral phenotypic state for *T. sirtalis*, and the genus *Thamnophis* in general, is high sensitivity to TTX, indicating that it is unlikely they could all have valine at this position, which produces a decrease in the binding affinity of the channel for TTX.

Our conclusions are based on analyses of differences in cDNA sequence that cannot be unequivocally attributed to changes in genomic DNA. A complete analysis of the relative importance of parallel genotypic evolution in this system will require much broader sampling of genomic DNA across a range of lineages, but the functional analyses of expressed proteins reported here indicate that both unique and parallel molecular changes might have a function in the diversification of TTX resistance in garter snakes. Although the changes in $\text{tsNa}_v1.4$ seem to explain much of the variation in TTX resistance among garter snake populations, considerable variation within populations has been observed that cannot be explained by the genetic mechanism reported here. These individual differences might be explained by alterations in the protein-coding regions of other genes or by differences in gene regulation, including that of $\text{tsNa}_v1.4$. Nonetheless, our study shows that adaptive diversification of a quantitative trait within a species proceeds through the variable effects of substitution of different amino acids at a highly conserved locus. □

Methods

cDNA library construction and screening

Total RNA was extracted from a skeletal muscle tissue sample from a 'Benton' snake (Benton County, Oregon; deposited in the University of Texas at Arlington Collection of Vertebrates, UTA R 52941) using TRIzol reagent as recommended by the manufacturer (Invitrogen). To construct the cDNA library, poly(A)⁺ RNA was isolated by Oligotex purification (Qiagen), then primed with random and oligo(dT) primers, and the cDNA was cloned into *EcoRI*-digested Lambda ZAP (Stratagene). The library was screened with random-primed, radiolabelled probes (Prime-It; Stratagene). One set of probes was developed from reverse transcriptase-mediated polymerase chain reaction (RT–PCR) products amplified from snake skeletal muscle tissue first-strand cDNA (SuperScript, Invitrogen) with the use of primer pairs designed from conserved $\text{Na}_v1.4$ and $\text{Na}_v1.5$

sequences (Supplementary Methods). All positive clones were sequenced with an ABI 373 or ABI 377 automated sequencer (Applied Biosystems) and aligned with Sequencer 3.1 software (Gene Codes Corp.). A $\text{Na}_v1.5$ -specific probe was prepared from RT–PCR products amplified from snake heart tissue first-strand cDNA (SuperScript) with the use of one primer pair (Supplementary Methods). No positive clones were identified with this probe.

RT–PCR

All other snake sequences were obtained from a skeletal muscle tissue sample of one individual from each population, 'Bear Lake' (Bear Lake County, Idaho), 'Warrenton' (Clatsop County, Oregon) and 'Willow Creek' (Sonoma County, California). Total RNA was extracted and cDNA was synthesized as above. PCR primers listed in Supplementary Methods were used to amplify the open-reading frame. RT–PCR products were cloned into pCR-Blunt (Invitrogen), sequenced and aligned.

Expression constructs

The full-length $\text{tsNa}_v1.4$ construct was prepared in pBluescript from three positive library clones. The *EcoRI* insert was ligated into pGH19 for expression in *Xenopus* oocytes. To prepare the human–snake chimaeric construct, regions within $\text{hNa}_v1.4$ /pSP64T were replaced with Benton $\text{tsNa}_v1.4$ by using a modified three-fragment PCR overlap extension: in domain I, $\text{tsNa}_v1.4$ residues 275–414 for $\text{hNa}_v1.4$ residues 277–424; in domain II, $\text{tsNa}_v1.4$ residues 761–814 for $\text{hNa}_v1.4$ residues 724–777; in domain III, $\text{tsNa}_v1.4$ residues 1,111–1,298 for $\text{hNa}_v1.4$ residues 1,185–1,269; and in domain IV, $\text{tsNa}_v1.4$ residues 1,537–1,603 for $\text{hNa}_v1.4$ residues 1,508–1,574. Other constructs were prepared by a two-fragment PCR overlap extension method of site-directed mutagenesis with substitutions at the positions shown in Fig. 1b. A second Willow Creek construct was prepared with a V1561I substitution. Amplified regions were verified by automated DNA sequencing.

Oocyte expression and electrophysiology

Capped RNAs were synthesized and purified (mMessage mMachine kit; Ambion). *Xenopus laevis* oocytes were collected and injected with 27 ng of each cRNA.

Ionic currents were measured at room temperature (22–25 °C) 2–10 days after cRNA injection by using the cut-open oocyte Vaseline gap voltage-clamp technique³⁰ with a CA-1B High Performance Oocyte Clamp (Dagan Instruments). Recordings were made in an external solution containing (in mM): 120 Na-Mes, 10 Hepes-Na, 1.8 CaCl_2 , pH 7.2, and an internal solution containing (in mM): 110 K-Mes, 10 Na-Mes, 10 Hepes-K, 1 EGTA, pH 7.2.

Current records were acquired using Pulse software (HEKA), sampling at 100 kHz and filtering at 20 kHz. Peak currents were evoked at 0.05 Hz with 20-ms pulses to 0 mV following a 200-ms prepulse to –120 mV. The holding potential for all experiments was –100 mV. Leak subtraction was performed before the test pulse (p) with the use of a $p/4$ protocol. Peak current amplitudes were measured offline with PulseFit (HEKA). The ratios of peak currents in the presence and absence of TTX over a range of TTX concentrations were calculated with peak currents recorded before and after perfusing the selected TTX concentration into the external bath solution for 5 min. To estimate the TTX concentration that blocked 50% of the expressed channels, the data were fitted to an equation derived from a single-site Langmuir adsorption isotherm: current ratio = $K_d/(K_d + [\text{TTX}])$, in which $[\text{TTX}]$ is the concentration of toxin and K_d is the concentration of TTX at which half of the receptors are bound to the toxin. Statistical analyses were conducted with PROC NLIN in SAS/STAT version 9.0. K_d and its 95% confidence limits were estimated from the curve.

Received 14 January; accepted 10 February 2005; doi:10.1038/nature03444.

1. Rausher, M. D., Miller, R. E. & Tiffin, P. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Mol. Biol. Evol.* **16**, 266–274 (1999).
2. Schluter, D., Clifford, E. A., Nemethy, M. & McKinnon, J. S. Parallel evolution and inheritance of quantitative traits. *Am. Nat.* **163**, 809–822 (2004).
3. Stern, D. L. Evolutionary developmental biology and the problem of variation. *Evolution* **54**, 1079–1091 (2000).
4. Brodie, E. D. Jr, Ridenhour, B. J. & Brodie, E. D. III The evolutionary response of predators to dangerous prey: hotspots and coldspots in the geographic mosaic of coevolution between garter snakes and newts. *Evolution* **56**, 2067–2082 (2002).
5. Beldade, P., Brakefield, P. M. & Long, A. D. Contribution of Distal-less to quantitative variation in butterfly eyespots. *Nature* **415**, 315–318 (2002).
6. Abzhonov, A., Protas, M., Grant, B. R., Grant, P. R. & Tabin, C. J. Bmp4 and morphological variation of beaks in Darwin's finches. *Science* **305**, 1462–1465 (2004).
7. Shapiro, M. D. et al. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**, 717–723 (2004).
8. Yamamoto, Y., Stock, D. W. & Jeffery, W. R. Hedgehog signalling controls eye degeneration in blind cavefish. *Nature* **431**, 844–847 (2004).
9. Hoekstra, H. E. & Nachman, M. W. Different genes underlie adaptive melanism in different populations of rock pocket mice. *Mol. Ecol.* **12**, 1185–1194 (2003).
10. ffrench-Constant, R. H., Daborn, P. J. & Le Goff, G. The genetics and genomics of insecticide resistance. *Trends Genet.* **20**, 163–170 (2004).
11. Hille, B. *Ion Channels of Excitable Membranes* (Sinauer, Sunderland, Massachusetts, 2001).
12. Brodie, E. D. III & Brodie, E. D. Jr Predator–prey arms races: asymmetrical selection on predators and prey may be reduced when prey are dangerous. *Bioscience* **49**, 557–568 (1999).
13. Janzen, F. J., Krenz, J. G., Haselkorn, T. S., Brodie, E. D. & Brodie, E. D. Molecular phylogeography of common garter snakes (*Thamnophis sirtalis*) in western North America: implications for regional historical forces. *Mol. Ecol.* **11**, 1739–1751 (2002).
14. Gefken, S., Brodie, E. D. Jr, Ruben, P. C. & Brodie, E. D. III Mechanisms of adaptation in a predator–prey arms race: TTX-resistant sodium channels. *Science* **297**, 1336–1339 (2002).

15. Satin, J. *et al.* A mutant of TTX-resistant cardiac sodium channels with TTX-sensitive properties. *Science* **256**, 1202–1205 (1992).
16. Kaneko, Y., Matsumoto, G. & Hanyu, Y. TTX resistivity of Na⁺ channel in newt retinal neuron. *Biochem. Biophys. Res. Commun.* **240**, 651–656 (1997).
17. Yotsu-Yamashita, M. *et al.* Binding properties of ³H-PbTx-3 and ³H-saxitoxin to brain membranes and to skeletal muscle membranes of puffer fish *Fugu pardalis* and the primary structure of a voltage-gated Na⁺ channel alpha-subunit (fMNa1) from skeletal muscle of *F. pardalis*. *Biochem. Biophys. Res. Commun.* **267**, 403–412 (2000).
18. Terlau, H. *et al.* Mapping the site of block by tetrodotoxin and saxitoxin of sodium channel II. *FEBS Lett.* **293**, 93–96 (1991).
19. Yamagishi, T., Li, R. A., Hsu, K., Marban, E. & Tomaselli, G. F. Molecular architecture of the voltage-dependent Na channel: functional evidence for alpha helices in the pore. *J. Gen. Phys.* **118**, 171–181 (2001).
20. Kallen, R. G. *et al.* Primary structure and expression of a sodium channel characteristic of denervated and immature rat skeletal muscle. *Neuron* **4**, 233–242 (1990).
21. Trimmer, J. S., Cooperman, S. S., Agnew, W. S. & Mandel, G. Regulation of muscle sodium channel transcripts during development and in response to denervation. *Dev. Biol.* **142**, 360–367 (1990).
22. Huey, R. B. & Moody, W. J. Neuroscience and evolution. Snake sodium channels resist TTX arrest. *Science* **297**, 1289–1290 (2002).
23. Zakon, H. H. Convergent evolution on the molecular level. *Brain Behav. Evol.* **59**, 250–261 (2002).
24. Lipkind, G. M. & Fozzard, H. A. KcsA crystal structure as framework for a molecular model of the Na⁺ channel pore. *Biochemistry* **39**, 8161–8170 (2000).
25. Plummer, N. W. & Meisler, M. H. Evolution and diversity of mammalian sodium channel genes. *Genomics* **57**, 323–331 (1999).
26. Raymond, C. K. *et al.* Expression of alternatively spliced sodium channel alpha-subunit genes: Unique splicing patterns are observed in dorsal root ganglia. *J. Biol. Chem.* **279**, 46234–46241 (2004).
27. Goldin, A. L. Resurgence of sodium channel research. *Annu. Rev. Physiol.* **63**, 871–894 (2001).
28. Choudhary, G., Yotsu-Yamashita, M., Shang, L., Yasumoto, T. & Dudley, S. C. Jr Interactions of the C-11 hydroxyl of tetrodotoxin with the sodium channel outer vestibule. *Biophys. J.* **84**, 287–294 (2003).
29. Penzotti, J. L., Fozzard, H. A., Lipkind, G. M. & Dudley, S. C. Jr Differences in saxitoxin and tetrodotoxin binding revealed by mutagenesis of the Na⁺ channel outer vestibule. *Biophys. J.* **75**, 2647–2657 (1998).
30. Stefani, E. & Bezanilla, F. Cut-open oocyte voltage-clamp technique. *Methods Enzymol.* **293**, 300–318 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Correa for advice regarding the cut-open oocyte voltage clamp; S. Durham for advice regarding the statistical analysis; C. Feldman and M. Pfrender for advice regarding the phylogenetic analysis; J. Caldwell for primers; A. Goldin for sharing his sodium channel sequence alignment; and C. Hanifin and the USU herpetology group for comments that improved the manuscript. This work was supported by research grants from the National Institute of Health (P.C.R.) and from the National Science Foundation (E.D.B. Jr and E.D.B. III).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.L.G. (geffeny@biology.usu.edu). Sequences described in this paper have been deposited in GenBank under accession numbers AY851743, AY851744, AY851745 and AY851746.

Sodium channel mutation leading to saxitoxin resistance in clams increases risk of PSP

V. Monica Bricelj¹, Laurie Connell², Keiichi Konoki³, Scott P. MacQuarrie¹, Todd Scheuer³, William A. Catterall³ & Vera L. Trainer⁴

¹Institute for Marine Biosciences, National Research Council, Halifax, Nova Scotia B3H 3Z1, Canada

²School of Marine Sciences, University of Maine, Orono, Maine 04469, USA

³Department of Pharmacology, University of Washington, Seattle, Washington 98195, USA

⁴NOAA Fisheries, Northwest Fisheries Science Center, Seattle, Washington 98112, USA

Bivalve molluscs, the primary vectors of paralytic shellfish poisoning (PSP) in humans, show marked inter-species variation in their capacity to accumulate PSP toxins (PSTs)¹ which has a neural basis^{2,3}. PSTs cause human fatalities by blocking sodium conductance in nerve fibres^{4,5}. Here we identify a molecular basis

for inter-population variation in PSP resistance within a species, consistent with genetic adaptation to PSTs. Softshell clams (*Mya arenaria*) from areas exposed to 'red tides' are more resistant to PSTs, as demonstrated by whole-nerve assays, and accumulate toxins at greater rates than sensitive clams from unexposed areas. PSTs lead to selective mortality of sensitive clams. Resistance is caused by natural mutation of a single amino acid residue, which causes a 1,000-fold decrease in affinity at the saxitoxin-binding site in the sodium channel pore of resistant, but not sensitive, clams. Thus PSTs might act as potent natural selection agents, leading to greater toxin resistance in clam populations and increased risk of PSP in humans. Furthermore, global expansion of PSP to previously unaffected coastal areas⁶ might result in long-term changes to communities and ecosystems.

PSP, caused by human consumption of shellfish that feed on toxic algae, is a public health hazard and causes severe economic losses globally due to bans on harvesting of contaminated shellfish and the need for costly monitoring programmes. PSP-producing dinoflagellates (for example *Alexandrium* spp.) cause toxic blooms ('red tides') in North America and worldwide. PSTs block conduction of the nerve impulse by interfering with the voltage-dependent increase in sodium-ion conductance that generates the action potential in nerve and muscle fibres^{4,5}, leading to neuromuscular paralysis. Large differences in PST accumulation between bivalve species *in vivo*¹ have been associated with *in vitro* differences in sensitivity of isolated nerves to saxitoxin (STX), the most potent PST, and the related tetrodotoxin (TTX)^{2,3}. STX and TTX bind to a single site in the outer pore of the Na⁺ channel, formed by the amino-acid residues in the outer-pore loops located between the S5 and S6 segments of each of the four homologous domains (I–IV) of the α -subunit^{7,8}. Could differences in amino acid sequence in the receptor site for STX and TTX in Na⁺ channels cause differences

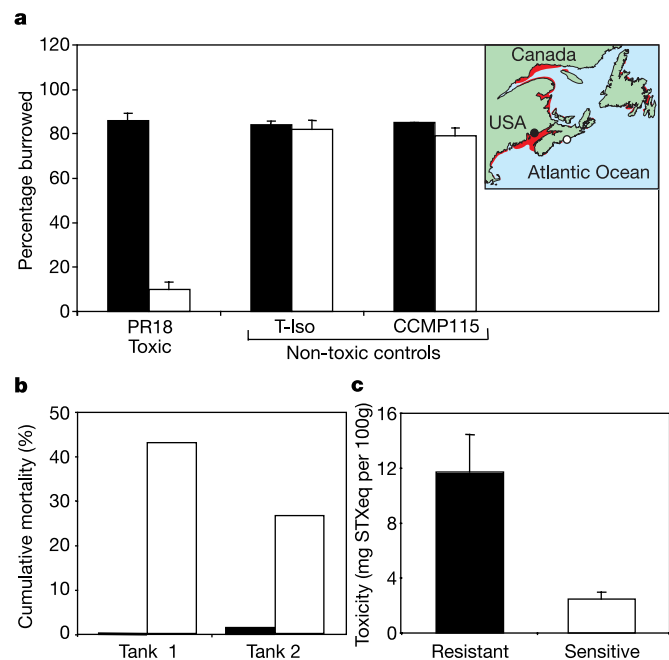


Figure 1 Responses to PSTs in two *M. arenaria* populations. **a**, Percentage of clams that burrowed after 24 h of exposure to *A. tamarensis* (strains PR18b or CCMP115) or *I. galbana* (T-Iso) ($n = 2$ tanks). Map shows the study sites BF (filled circle) and LE (open circle); the PSP-affected coastline is in red. **b**, Mortality after 16 days of toxification with strain PR18b (determined by visual inspection (dead clams removed) (tank 1) and by removal of all clams every 2 days (live clams reintroduced in sediment) (tank 2)). In **a** and **b**, filled columns are results for BF clams, and open columns for LE clams. **c**, Tissue toxicity of live clams after 16 days of toxification of resistant BF and sensitive LE clams (burrowers and non-burrowers at 24-h, respectively). All results are means $\pm$ s.e.m.

in toxin sensitivity and accumulation in shellfish and thereby contribute to the risk of PSP? We have probed this question with a combination of behavioural, physiological and molecular biological approaches.

Mya arenaria is a commercially important native species with wide latitudinal distribution in Atlantic North America, from the Gulf of St Lawrence to Chesapeake Bay. We collected clams from two sites with contrasting histories of PSP (Fig. 1a inset): Lepreau Basin, Bay of Fundy (BF), where annual, recurrent toxic blooms of *Alexandrium* spp. occur in summer⁹, and the Lawrencetown estuary (LE), Nova Scotia, an area with no record of PSP. Burrowing activity provided an index of susceptibility to toxins in clams¹⁰. After exposure to toxic *A. tamarens* cells for 24 h, most clams (a mean of 86%) from LE were unable to re-burrow after deployment at the sediment surface, whereas only 10% of clams from the BF population were compromised (Fig. 1a). Burrowing was not affected by exposure to non-toxic *Isochrysis galbana* (T-Iso) or to *A. tamarens* strain CCMP115, which is of negligible toxicity. Therefore burrowing incapacitation in susceptible clams is toxin-induced and is attributed to muscle paralysis. PSTs also prevented siphon retraction in susceptible clams¹¹. Sublethal effects of PSTs on burrowing and siphon retraction in a high-energy, intertidal habitat, where clams can become exposed at the sediment surface, may cause indirect mortalities through desiccation and predation.

Longer-term laboratory exposure to toxic *A. tamarens* cells resulted in high differential mortalities between the two populations (Fig. 1b). Mortalities of LE clams were consistently higher (26–42% after 16 days of toxin exposure) than those of BF clams held under identical conditions (2% or less). LE clams also showed significantly lower feeding rates on toxic cells and decreased metabolic rates during toxification¹¹. This individual variation in fitness-related traits provides a basis for natural selection.

To compare their toxin accumulation capacity, dominant phenotypes from the two populations were exposed to toxic *A. tamarens* for 16 days. Resistant BF clams (burrowers after 24 h of toxification) attained a mean toxicity significantly (fivefold) higher than sensitive LE clams (non-burrowers) (Fig. 1c, equal to a toxin concentration of 657 and 112 nmol g⁻¹ respectively). Toxicities in both groups exceeded the regulatory safety level (80 µg STX equivalents per 100 g) and increased linearly at rates of 771 ($r^2 = 0.96$) and 113 µg STX equivalents per day ($r^2 = 0.80$) in resistant and sensitive clams, respectively (data not shown), resulting in more than an order of magnitude difference in toxicity among individuals after 2 weeks. *M. arenaria* field populations can attain toxicities (~9,600 µg STX equivalents per 100 g) comparable to those determined in this study and accumulate even higher toxin levels in the viscera¹.

In vitro exposure of isolated cerebrovisceral nerve trunks of naive (toxin-free) clams from BF and LE to serially increasing STX concentrations also revealed large (more than 100-fold) differences in sensitivity to STX between and within populations (Fig. 2a; compare results at 4 and 400 µM STX). Nerves from most clams in the LE population with no history of PSP (69%) experienced full block of the action potential at 33 µM STX within ~30 s (Fig. 2b, left). In contrast, most clams (91%) from the PSP-exposed population (BF) were resistant to 33 µM and even 334 µM STX, the highest concentration tested. These differences in response in naive clams indicate that they were not induced by toxins retained in tissues but were an intrinsic response of individual nerves to toxin exposure *in vitro*.

Site-directed mutations within the Na⁺ channel pore region¹² and naturally occurring variations in amino acid sequences in the outer pores of different mammalian Na⁺ channels^{13,14} can result in a decreased binding affinity for STX and TTX and thus in an increased toxin resistance. To test the hypothesis that a mutation in the α -subunit Na⁺ channel is responsible for the differences in STX sensitivity, we determined both the genomic and the complementary DNA sequence of the pore regions of Na⁺ channels

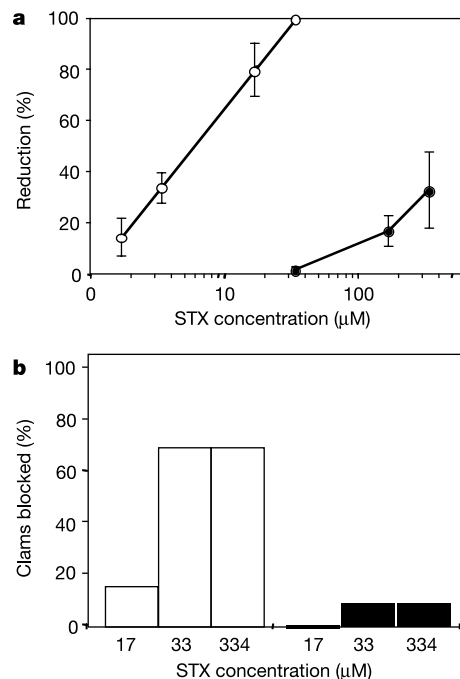


Figure 2 Nerve response of toxin-free, individual clams to STX *in vitro*. **a**, Percentage decrease in the compound action potential in isolated nerve preparations exposed to increasing STX concentrations relative to the pre-exposure condition. Values are amplitudes of the action potential of three representative sensitive (open circles) and three resistant (filled circles) clams (means $\pm$ s.e.m.). The response measured when the reduction was maximized (~5 min except for the 100% block, which occurs within seconds). **b**, Percentage of clams tested from each population exhibiting full action potential block in relation to STX concentration (the maximum tested was 334 µM; $n = 11$ BF clams (filled columns) and 13 LE clams (open columns)).

from sensitive and resistant *Mya arenaria* individuals, identified by using the nerve assay. Alignments of the predicted protein sequences encoded by the cDNAs showed striking similarities to known voltage-gated Na⁺ channels from jellyfish to rat (Fig. 3). Alignments of all individual clam pore region sequences revealed a single-nucleotide mutation that correlated with clam resistance to STX. The genomic DNA and cDNA sequences were identical for each individual clam, showing that RNA editing¹⁵ was not involved in generation of the mutated Na⁺ channel protein. This mutation resulted in an amino-acid substitution in the outer pore loop of domain II from glutamic acid (E) to aspartic acid (D) at a position equivalent to E945 in rat Nav1.2a (ref. 16). Glutamic acid is found at this position in all wild-type (WT) Na⁺-channel α subunits from a wide range of species (Fig. 3). Molecular models propose that this residue forms part of the outer pore lining and interacts with the hydroxyl groups at C-12 of STX¹⁷. Site-directed *in vitro* mutagenesis experiments have shown that it is a major determinant for STX binding¹². Neutralization of this acidic residue by mutation to glutamine (Q) decreased STX affinity 19,880-fold and TTX affinity 290-fold compared with WT¹⁸.

To assess the role of the E945D mutation in the block by STX *in vitro*, the mutation was introduced by means of a shuttle plasmid into the corresponding position of rat brain Nav1.2a channels. The mutant Na⁺ channel produced currents with electrophysiological properties indistinguishable from those of the WT channel (Fig. 4a, b, and Supplementary Fig. S1). Application of 10 nM TTX blocked about half of the WT current, and 100 nM TTX blocked it almost completely (Fig. 4a, top). The half-maximal inhibitory concentration (IC₅₀) for WT was 11.2 ± 1.5 nM. In contrast, 1,000-fold more TTX (10 µM) blocked only ~20% of the current through E945D channels (Fig. 4a, bottom) and the IC₅₀ increased about

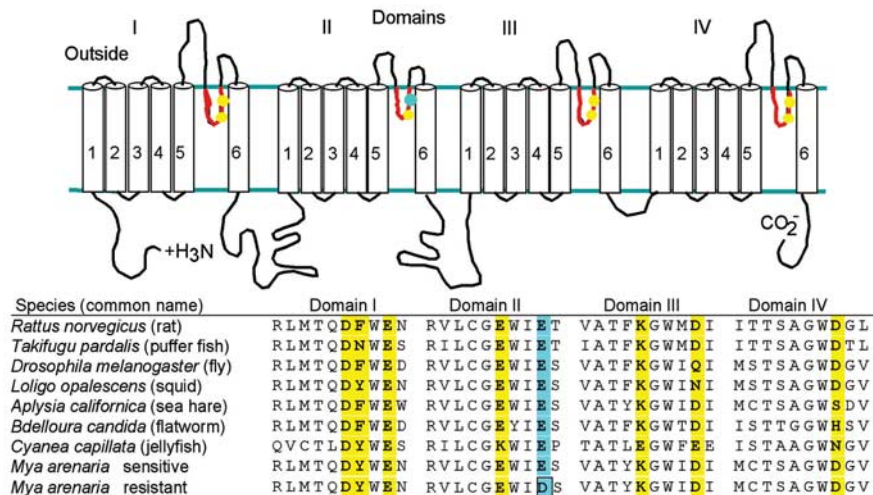


Figure 3 Mutation in the Na⁺ channel pore of resistant *M. arenaria*. Top, primary structure of the Na⁺ channel α -subunit, shown as a two-dimensional protein folding in and out of the cell membrane⁸. Cylinders represent transmembrane α -helices; solid lines represent hydrophilic portions of the sequence. Amino acids of the pore region (SS1 and SS2) in each domain are shown in red, and those associated with STX binding in yellow.

Bottom, alignments of pore-forming loops (SS1, SS2) in each of four Na⁺-channel domains in *M. arenaria* compared with those in other organisms, in descending phylogenetic order (GenBank database). Highlighted residues are crucial to STX binding^{12,13}; position 945 (blue) in resistant *M. arenaria* is boxed.

3,000-fold to $35 \pm 9 \mu\text{M}$ (Fig. 4c). Because STX is more potent, an order of magnitude less toxin was required to obtain a similar blocking effect. STX at 1 nM blocked 40% of WT current, and 10 nM STX blocked it nearly completely (Fig. 4b, top) ($\text{IC}_{50} = 1.7 \pm 0.2 \text{ nM}$). In contrast, a 1,000-fold higher STX concentration (1 μM) blocked only 25% of the E945D current (Fig. 4b, bottom) ($\text{IC}_{50} \approx 2.7 \pm 0.6 \mu\text{M}$; Fig. 4d). Thus, the E945D substitution made the Nav1.2a channel about 1,500-fold or 3,000-fold less sensitive to STX and TTX, respectively, and a similar decrease in affinity is expected in the *M. arenaria* Na⁺ channel (see Supplemen-

tary Discussion). This is a surprisingly large reduction in binding affinity resulting from simply shortening of the side chain at position 945 by one methylene group without changing the negative charge of the carboxyl group. These results imply that this side chain is rigid in the outer vestibule of the channel and emphasize that even a small conformational change in the receptor site can result in a large decrease in the binding affinity of these toxins to Na⁺ channels by altering the spatial relationships between ion-pair-forming and hydrogen-bond-forming partners.

Our results show that marked differences in whole-animal and

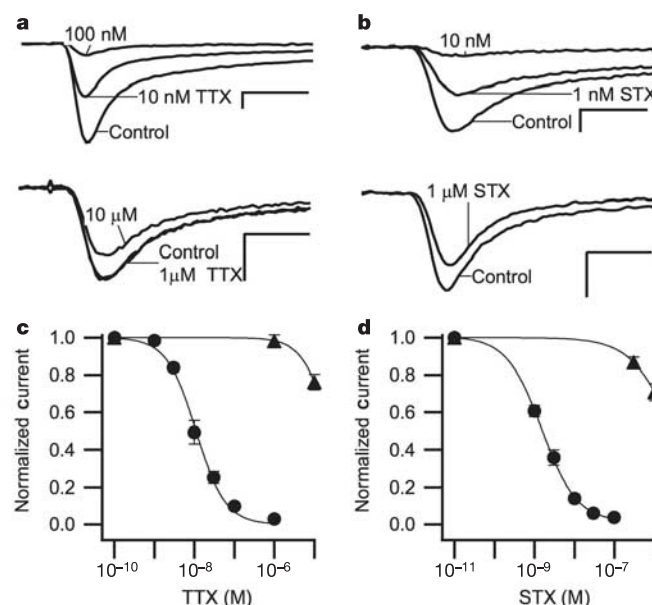


Figure 4 Blocking of WT and mutant Na⁺ channels by TTX and STX. **a, b**, Na⁺ currents in tsA-201 cells expressing WT (top) and E945D (bottom) channels and their blocking by TTX (**a**) and STX (**b**). Currents elicited by step depolarizations to -10 mV from a -120-mV holding potential. Scale bars, 1 nA and 1 ms. **c, d**, Inhibition curves for block of WT

(circles) and E945D (triangles) by TTX (**c**) and STX (**d**). Peak Na⁺ current normalized to its control is plotted against toxin concentration. Values are means $\pm$ s.d. for at least three determinations. Smooth curves were fitted with Langmuir isotherms to evaluate IC_{50} .

nerve susceptibility to PSTs and in toxin uptake capacity in *M. arenaria* result from a single mutation in domain II of the Na⁺ channel pore region. These observed differences correspond to inter-population differences in the history of exposure to toxic blooms in the environment. The adaptive value of resistance to PSTs (increased survival) was also shown experimentally. Our results support the conclusion that *M. arenaria* populations in PSP-affected areas undergo genetic adaptation to toxins through selective mortality or reduced fitness of sensitive individuals. In this respect *M. arenaria* differ from Pacific butter clams, *Saxidomus giganteus*, in which PSTs function as an anti-predator defence¹⁹.

A neural basis for inherited toxin resistance was shown in cattle ticks and house flies, in which a single Na⁺ channel mutation confers resistance to pesticides (pyrethroids and DDT)^{20,21}. Resistance to TTX in puffer fish *Takifugu (Fugu) pardalis*, which can accumulate high concentrations of TTX in tissue without adverse effects, has been attributed to a single mutation in the skeletal-muscle Na⁺ channel²², yet natural selection for resistance was not shown in this species. In terrestrial ecosystems, marked geographic variation in resistance to TTX in populations of garter snakes (*Thamnophis sirtalis*) has coevolved with that of its prey²³ (the toxic newt, *Taricha* sp.). Thus, different selective pressures (predation or toxic blooms) can lead to geographic differentiation in resistance to natural toxins within species. The increased accumulation of toxin in resistant *M. arenaria* points to this resistance mutation as an important risk factor for human PSP resulting from the consumption of this species. Our findings raise the possibility that other bivalve species might harbour similar mutations, thus allowing further understanding of the molecular basis of toxin resistance across shellfish species. □

Methods

Whole-animal observations

Juvenile clams (~30–47 mm shell height), collected at a time of year when they contained no PSTs, were acclimated (16 °C, 30‰ salinity) for at least 2–3 weeks before experiments. Clams were toxicified, as described previously²⁴, in aquaria containing coarse sand 10 cm deep, at a constant (100 cells ml⁻¹) bloom concentration of *A. tamarensis* (isolate PR18b from the Gulf of St Lawrence; toxicity 60–98 pg STX equivalents per cell), maintained by cell delivery with a peristaltic pump¹¹. Burrowing response was measured with previously described protocols¹⁰ after exposure of clams to equal biovolume concentrations of either toxic or non-toxic cells.

For the measurement of survival, 40–50 clams from each population (BF and LE) were held together in each of two experimental tanks exposed to *A. tamarensis* (PR18b). Mortalities were not averaged (Fig. 1b)—that is, tanks were not treated as replicates—because mortality determinations differed between the two systems (see the legend to Fig. 1b). Because of the lack of responsiveness of live but incapacitated (paralysed) clams, only clams exhibiting obvious signs of death (blackened and decomposing tissues, odour) were removed and included in mortality counts. Clams from both populations exposed in parallel to non-toxic *Thalassiosira weissflogii* suffered no mortalities over the course of the experiment.

To determine toxin accumulation, initially toxin-free clams were preselected (100 resistant BF clams and 120 sensitive LE clams based on the 24-h burrowing index) and then subjected to continued toxicification in a common aquarium. The toxicity of algae and individual clams (five resistant BF and five sensitive LE clams) sampled every 2 days was determined by following methods described previously^{11,24}. Individual PSTs were analysed by reverse-phase ion-pair high-performance liquid chromatography (HPLC) with fluorescence detection and converted to STX equivalents²⁵. Toxin standards were obtained from the Institute of Marine Biosciences' Certified Reference Materials Program (CRMP). Final toxicities were compared by the analysis of variance and linear regressions fitted to toxicity data over time. The initial toxin-free status of experimental clams in all trials was confirmed by HPLC ($n = 3–5$).

Nerve electrophysiology was examined by the nerve test, adapted from Twarog^{2,3}, which was conducted at 14–15 °C using sections of nerve trunk no larger than 2 cm, without removal of the neural sheath. STX standard (3.48 mM in 0.1 M acetic acid) was diluted to required test concentrations (1.7–334 µM STX) in physiological saline solution (PSS) containing (in mM) 428 NaCl, 10 KCl, 20 MgCl₂·6H₂O, 10 CaCl₂·2H₂O and 50 Tris-HCl, adjusted to pH 7.4 with 1.0 M HCl and bubbled with air. Control tests indicated that the nerve response was not affected by acetic acid at the dilutions used (tested at constant pH) or by a pH of 5.6 or less. Thus no pH adjustment of the PSS was necessary for tests at 334 µM STX or less (the pH of the test solution was 6.2 to 7.4). Clams that showed a full action potential block at no more than 33 µM STX were characterized as sensitive, and those that required at least 334 µM STX to fully block were characterized as resistant.

Molecular genetics

For determination of the Na⁺ channel sequence, nerve connectives dissected from living *M. arenaria* individuals were used for nerve assays, and the ganglia from the same individual were immediately transferred to RNAlater (Ambion) for storage. RNA was extracted from individual clam nerve tissue and stored in RNasecure (Ambion) at –80 °C.

First-strand cDNA was produced from isolated DNA-free RNA with the use of SuperScript II reverse transcriptase (Invitrogen) and random hexamer primers, followed by digestion with RNase H in accordance with the manufacturer's instructions. *M. arenaria*-specific Na⁺-channel fragments were generated from the first-strand template by nested polymerase chain reaction (PCR) starting with domain IV and walking towards domain I. The complete cDNA sequence was determined for a fragment spanning domains I–IV (4,034 base pairs) using Big Dye Terminator Cycle Sequencing in an ABI 377 autosequencer (Applied Biosystems).

DNA and RNA were isolated from four individual clams from each site that had previously been subjected to nerve assays. Sequence from both the DNA and the resulting cDNA was obtained for each of the pore regions by using *M. arenaria*-specific PCR primers. The resulting PCR products were amplified and sequenced.

For the construction of Na⁺ channel mutations, a two-step mutagenesis protocol, modified from previous methods²⁶, was followed to generate the specific E945D mutation through the shuttle vector pBluescript SK+/rNav 1.2a (*Xma*I > *Bst*EII) that contains the *Xma*I (nucleotide 1,896 in the loop between DI and II) to *Bst*EII (nucleotide 4,636 in domain IV region S1) into a fragment of full-length rat pCDM8/ Nav1.2a (rIIa) through the *Sph*I and *Bgl*II restriction sites.

Further details are given in Supplementary Methods.

Electrophysiology of transfected Na⁺ channels

Human tsA-201 cells were transfected with cDNAs encoding WT or E945D mutant Nav1.2a channels by calcium phosphate precipitation, and Na⁺ currents were recorded by whole-cell voltage clamp as described previously²⁷. The extracellular solution contained (in mM) 140 NaCl, 5.4 CsCl, 1.8 CaCl₂, 1 MgCl₂, 10 Hepes (pH 7.35) at ~23 °C. The intracellular saline contained (in mM) 189 N-methyl-D-glucamine (NMDG), 1 NaCl, 4 MgCl₂, 0.1 1,2-bis(2-aminophenoxy)ethane-*N,N,N',N'*-tetraacetic acid (BAPTA), 25 Tris-phosphocreatine, 2 NaATP, 0.2 NaGTP and 40 Hepes, pH 7.35. TTX was obtained from Calbiochem, and STX standard (480 ± 20 µM STX in 0.1 M acetic acid) was from CRMP.

Received 13 December 2004; accepted 4 February 2005; doi:10.1038/nature03415.

1. Bricelj, V. M. & Shumway, S. E. Paralytic shellfish toxins in bivalve mollusks: occurrence, transfer kinetics, and biotransformation. *Rev. Fish. Sci.* **6**, 315–383 (1998).
2. Twarog, B. M., Hidaka, T. & Yamaguchi, H. Resistance to tetrodotoxin and saxitoxin in nerves of bivalve mollusks. *Toxicon* **10**, 273–278 (1972).
3. Twarog, B. M. in *Proc. 2nd Int. Coral Reef Symp.* Vol. 1 (eds Cameron, A. M. et al.) 505–512 (Barrier Reef Committee, Brisbane, 1974).
4. Narahashi, T. & Moore, J. W. Neuroactive agents and nerve membrane conductances. *J. Gen. Physiol.* **51**, 93–101 (1968).
5. Hille, B. Pharmacological modifications of the sodium channels of frog nerve. *J. Gen. Physiol.* **51**, 199–219 (1968).
6. Hallegraef, G. M. in *Manual on Harmful Marine Microalgae* (eds Hallegraef, G. M., Anderson, D. M. & Cembella, A. D.) 25–49 (UNESCO, Paris, 2003).
7. Fozzard, H. A. & Hanck, D. Structure and function of voltage-dependent sodium channels: comparison of brain II and cardiac isoforms. *Physiol. Rev.* **76**, 887–926 (1996).
8. Catterall, W. A. From ionic currents to molecular mechanisms: the structure and function of voltage-gated sodium channels. *Neuron* **26**, 13–25 (2000).
9. Martin, J. L. & Richard, D. in *Harmful and Toxic Algal Blooms* (eds Yasumoto, T., Oshima, Y. & Fukuyo, Y.) 3–6 (International Oceanographic Commission of UNESCO, Paris, 1996).
10. Bricelj, V. M., Cembella, A. D., Laby, D., Shumway, S. E. & Cucci, T. L. in *Harmful and Toxic Algal Blooms* (eds Yasumoto, T., Oshima, Y. & Fukuyo, Y.) 405–408 (International Oceanographic Commission of UNESCO, Paris, 1996).
11. MacQuarrie, S. P. *Inter- and Intra-population Variability in Behavioral and Physiological Responses of the Softshell Clam, Mya arenaria, to the PSP Toxin-producing Dinoflagellate, Alexandrium tamarensis*. Thesis, Dalhousie Univ. (2002).
12. Terlau, S. H. et al. Mapping the site of block by tetrodotoxin and saxitoxin of sodium channel II. *FEBS Lett.* **293**, 93–96 (1991).
13. Satin, J. et al. A mutant of TTX-resistant cardiac sodium channels with TTX-sensitive properties. *Science* **256**, 1202–1205 (1992).
14. Sivillotti, L., Okuse, K., Akopian, A. N., Moss, S. & Wood, J. N. A single serine residue confers tetrodotoxin insensitivity on the rat sensory-neuron-specific sodium channel SNS. *FEBS Lett.* **409**, 49–52 (1997).
15. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genet.* **33** (Suppl.), 245–254 (2003).
16. Auld, V. J. et al. A neutral amino acid change in segment IIS4 dramatically alters the gating properties of the voltage-dependent sodium channel. *Proc. Natl Acad. Sci. USA* **87**, 323–327 (1990).
17. Lipkind, G. & Fozzard, H. A. A structural model of the tetrodotoxin and saxitoxin binding site of the Na⁺ channel. *Biophys. J.* **66**, 1–13 (1994).
18. Kontis, K. J. & Goldin, A. L. Site-directed mutagenesis of the putative pore region of the rat IIA sodium channel. *Mol. Pharmacol.* **43**, 635–644 (1993).
19. Kvitek, R. G. & Beitler, M. K. Relative insensitivity of butter clam neurons to saxitoxin: a pre-adaptation for sequestering paralytic shellfish poisoning toxins as a chemical defense. *Mar. Ecol. Prog. Ser.* **69**, 47–54 (1991).
20. Liu, M.-Y., Bull, D. L. & Plapp, F. W. Jr Effects of exposure to cypermethrin on saxitoxin binding in susceptible and pyrethroid-resistant houseflies. *Arch. Insect Biochem. Physiol.* **37**, 73–79 (1998).
21. He, H. et al. Identification of a point mutation in the para-type sodium channel gene from a pyrethroid-resistant cattle tick. *Biochem. Biophys. Res. Commun.* **261**, 558–561 (1999).
22. Yamashita, M.-Y. et al. Binding properties of ³H-PbTx-3 and ³H-saxitoxin to brain membranes and to

skeletal muscle membranes of puffer fish *Fugu pardalis* and the primary structure of a voltage-gated Na^+ channel α -subunit (fMNa1) from skeletal muscle of *F. pardalis*. *Biochem. Biophys. Res. Commun.* **267**, 403–412 (2000).

23. Geffeney, S., Brodie, E. D. Jr, Ruben, P. C. & Brodie, E. D. III Mechanisms of adaptation in a predator-prey arms race: TTX-resistant sodium channels. *Science* **297**, 1336–1339 (2002).
24. Bricej, V. M., Lee, J. H. & Cembella, A. D. Influence of dinoflagellate cell toxicity on uptake and loss of paralytic shellfish toxins in the northern quahog, *Mercenaria mercenaria* (L.). *Mar. Ecol. Prog. Ser.* **74**, 33–46 (1991).
25. Oshima, Y. in *Manual on Harmful Marine Microalgae* (eds Hallegraeff, G. M., Anderson, D. M. & Cembella, A. D.) 81–94 (International Oceanographic Commission Manuals and Guides 33, UNESCO, Paris, 1995).
26. Storic, L., Lewis, K. & Resnick, M. A. *In-vivo* site-directed mutagenesis using oligonucleotides. *Nature Biotechnol.* **19**, 773–776 (2001).
27. Linford, N. J., Cantrell, A. R., Qu, Y., Scheuer, T. & Catterall, W. A. Interaction of batrachotoxin with the local anesthetic receptor site in transmembrane segment IVS6 of the voltage-gated sodium channel. *Proc. Natl Acad. Sci. USA* **95**, 13947–13952 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank B. M. Twarog, whose seminal work in the 1970s inspired this study, for conducting the initial nerve tests; P. Chang for participating in the burrowing experiment; M. Quilliam and the IMB analytical toxins group for providing STX for nerve tests; and E. M. Sharp and M. Iszard for technical assistance. This work was supported by a US NOAA-ECOHAB grant to V.L.T. and V.M.B., a NOAA-ECOHAB grant to L.C., V.L.T. and V.M.B., and an NIH research grant to W.A.C.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to V.M.B. (monica.bricej@nrc-cnrc.gc.ca). The complete sequence of the Na^+ channel pore region has been submitted to the GenBank database under accession no. AY847740.

Ipr1 gene mediates innate immunity to tuberculosis

Hui Pan^{1*}, Bo-Shiun Yan^{1*}, Mauricio Rojas^{1,3*}, Yuriy V. Shebzukhov^{1†}, Hongwei Zhou², Lester Kobzik², Darren E. Higgins⁴, Mark J. Daly⁵, Barry R. Bloom¹ & Igor Kramnik¹

¹Department of Immunology and Infectious Diseases and

²Physiology Program, Department of Environmental Health, Harvard School of Public Health, 667 Huntington Avenue, Boston, Massachusetts 02115, USA

³Grupo de Inmunología Celular e Inmunogenética, Facultad de Medicina, Universidad de Antioquia, Medellín, Colombia

⁴Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

⁵Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA

* These authors contributed equally to this work

† Present address: Department of Molecular Immunology, A. N. Belozersky Institute of Physical and Chemical Biology, Moscow State University, Vorobjov Gory, Moscow, 119899, Russia

An estimated eight million people are infected each year with the pathogen *Mycobacterium tuberculosis*, and more than two million die annually¹. Yet only about 10% of those infected develop tuberculosis. Genetic variation within host populations is known to be significant in humans and animals^{2,3}, but the nature of genetic control of host resistance to tuberculosis remains poorly understood. Previously we mapped a new genetic locus on mouse chromosome 1, designated *sst1* (for super-susceptibility to tuberculosis 1)⁴. Here we show that this locus mediates innate immunity in *sst1* congenic mouse strains and identify a candidate gene, *Intracellular pathogen resistance 1* (*Ipr1*), within the *sst1* locus. The *Ipr1* gene is upregulated in the *sst1* resistant macrophages after activation and infection, but it is not expressed in the *sst1* susceptible macrophages. Expression of the *Ipr1* transgene in the *sst1* susceptible macrophages limits the multiplication not only of *M. tuberculosis* but

also of *Listeria monocytogenes* and switches a cell death pathway of the infected macrophages from necrosis to apoptosis. Our data indicate that the *Ipr1* gene product might have a previously undocumented function in integrating signals generated by intracellular pathogens with mechanisms controlling innate immunity, cell death and pathogenesis.

It is estimated that about one-third of the human population on the planet has been infected by virulent *M. tuberculosis*^{1,5}. Susceptibility to clinical tuberculosis is known to be influenced by environmental factors such as stress, malnutrition, concomitant infections (for example HIV) or senescence^{6,7}. Although genetic variation within host populations is also known to affect resistance and susceptibility, individual genes responsible for innate immunity to the pathogen have been elusive. In susceptible individuals, progression of lung tuberculosis often leads to the formation of characteristic necrotic 'cavities' that destroy significant portions of the lung. Beyond their life-threatening clinical consequences, these lesions are essential for the efficient transmission of *M. tuberculosis* in aerosols. Because tuberculosis in humans is transmitted primarily by the respiratory route, the ability to cause lung disease is considered a key aspect of the pathogen's virulence strategy and ensures its evolutionary success. Therefore, understanding pathogenic mechanisms that are employed by virulent *M. tuberculosis* during lung tuberculosis in susceptible individuals is essential for developing effective prevention and treatment strategies^{8,9}. However, detailed mechanistic studies of pathogenesis of lung tuberculosis and its genetic control have been limited by the fact that in mouse models of *M. tuberculosis* infection, necrotic lesions in the lungs are rarely found unless the mouse is rendered systemically immunodeficient.

C3HeB/FeJ inbred mice are extremely susceptible to virulent *M. tuberculosis* and develop a marked lung pathology, which leads to their rapid death after infection^{10,11}. We generated a congenic mouse strain C3H.B6-*sst1* (*sst1*^R) carrying the C57BL/6J-derived resistant allele at the *sst1* locus on the C3HeB/FeJ genetic background. The survival time of the *sst1*^R congenic mice infected either with a high dose of intravenous *M. tuberculosis* (Fig. 1a) or with a low dose of *M. tuberculosis* by the respiratory route (Fig. 1b), relative to their *sst1*^S counterparts, is significantly lengthened, indicating a profound effect of the locus on anti-tuberculosis immunity. However, the shorter survival of the C3H.B6-*sst1* (*sst1*^R) mice, in comparison with the resistant parental strain C57BL/6J (B6), indicates that the *sst1* locus is responsible for a significant portion, but not all, of the tuberculosis resistance phenotype of the B6 mice.

The specific effect of the *sst1* locus on the progression of tuberculosis was related to a more efficient control of *M. tuberculosis* multiplication, primarily in the lungs, after both respiratory challenge by aerosol (Fig. 1c) and systemic intravenous infection (Supplementary Fig. 2a). The development of large necrotic lung lesions within 4 weeks after intravenous infection, characteristic of *sst1*^S mice, was prevented in the presence of the *sst1*^R allele (Fig. 1d). After a low-dose aerosol infection, chronic tuberculosis infection ensued, and the *sst1*^S mice developed encapsulated necrotic lung lesions, in some cases reaching about one-third of the lung lobe (Fig. 1e), that resembled tuberculosis cavities in human lungs. Mycobacteria were present both extracellularly, within necrotic central areas surrounded by the fibrotic capsule, and within macrophages of the granuloma wall (Supplementary Fig. 1). In the *sst1*^R mice, lung lesions were much smaller and contained fewer infected macrophages.

Although the greatest effect of the *sst1* polymorphism on the progression of tuberculosis was observed in the lungs, bone marrow transplantation experiments showed that bone marrow-derived cells, but not lung cells, were responsible for the effect of the *sst1* locus (Supplementary Fig. 2b). It is known that T lymphocytes and macrophages are of major importance in host resistance to tuberculosis. We have found that, whereas T lymphocytes are functionally

unaffected by the *sst1* polymorphism, the *sst1* disparate macrophages show considerable differences in their ability to control *M. tuberculosis* *in vitro* (M.R., J. Gutierrez-Pabello, H.P., O. Jobe, B.-S.Y., L. Helming, L.K. and I.K., manuscript in preparation). The rate of *M. tuberculosis* multiplication was significantly higher in the *sst1^S* macrophages (Fig. 1g, left panel). There was also a clear distinction in the mechanism of macrophage cell death after the infection: the *sst1^S* macrophages showed characteristic necrosis, whereas the *sst1^R* macrophages underwent apoptosis (Fig. 1f, upper left and upper right panels, respectively). The effect of the *sst1* locus was much more pronounced after infection of macrophages with virulent *M. tuberculosis*, because an avirulent vaccine strain of *M. bovis* bacillus Calmette–Guérin (BCG) failed to multiply in the *sst1^S* macrophages (Fig. 1g, right panel) and to induce necrosis of them (Fig. 1f, lower left panel).

In vivo, many cells within the tuberculosis lung granulomas of the *sst1^R* mice contained apoptotic nuclei that were positive under TdT-mediated dUTP nick end labelling (TUNEL assay) and no necrosis was observed, whereas the apoptotic nuclei were largely absent from the necrotic lesions of the *sst1^S* mice (see Supplementary Information). Thus our studies, both *in vivo* and *in vitro*, indicated that the extreme susceptibility to virulent *M. tuberculosis*

of *sst1^S* mice is associated with necrotic death of the susceptible macrophages. Virulent *M. tuberculosis* was shown to cause necrosis of the infected epithelial cell lines¹². This necrosis-inducing propensity is specific for virulent *M. tuberculosis* and is lost in RD1 mutants of the microbe, which also have a markedly decreased virulence^{12,13}. From our studies it seems that, in addition to virulence determinants of the pathogen itself, mechanisms of host cell death depend on the host polymorphic gene(s) encoded within the *sst1* locus.

To identify the critical gene(s), we employed a positional cloning strategy (see Supplementary Information and Supplementary Fig. 3 for details). First, the *sst1* minimal region was reduced to an interval between D1Mit439 and D1Mit49 markers on mouse chromosome 1 (Supplementary Fig. 3a). This region contains a so-called HSR (for homogeneously stained region) repeat (Fig. 2a). The HSR repeat region is arguably the largest repetitive region in the mouse genome^{14,15}. Its size in inbred mouse strains is estimated to be between 3.5 million and 6 million bases (refs 15, 16) and it remains unfinished by both mouse genome projects. After identifying and testing progeny of additional recombinants within this interval (Fig. 2a), we concluded that the *sst1* candidate region encompasses part of the HSR repeat region and a region of mouse chromosome 1

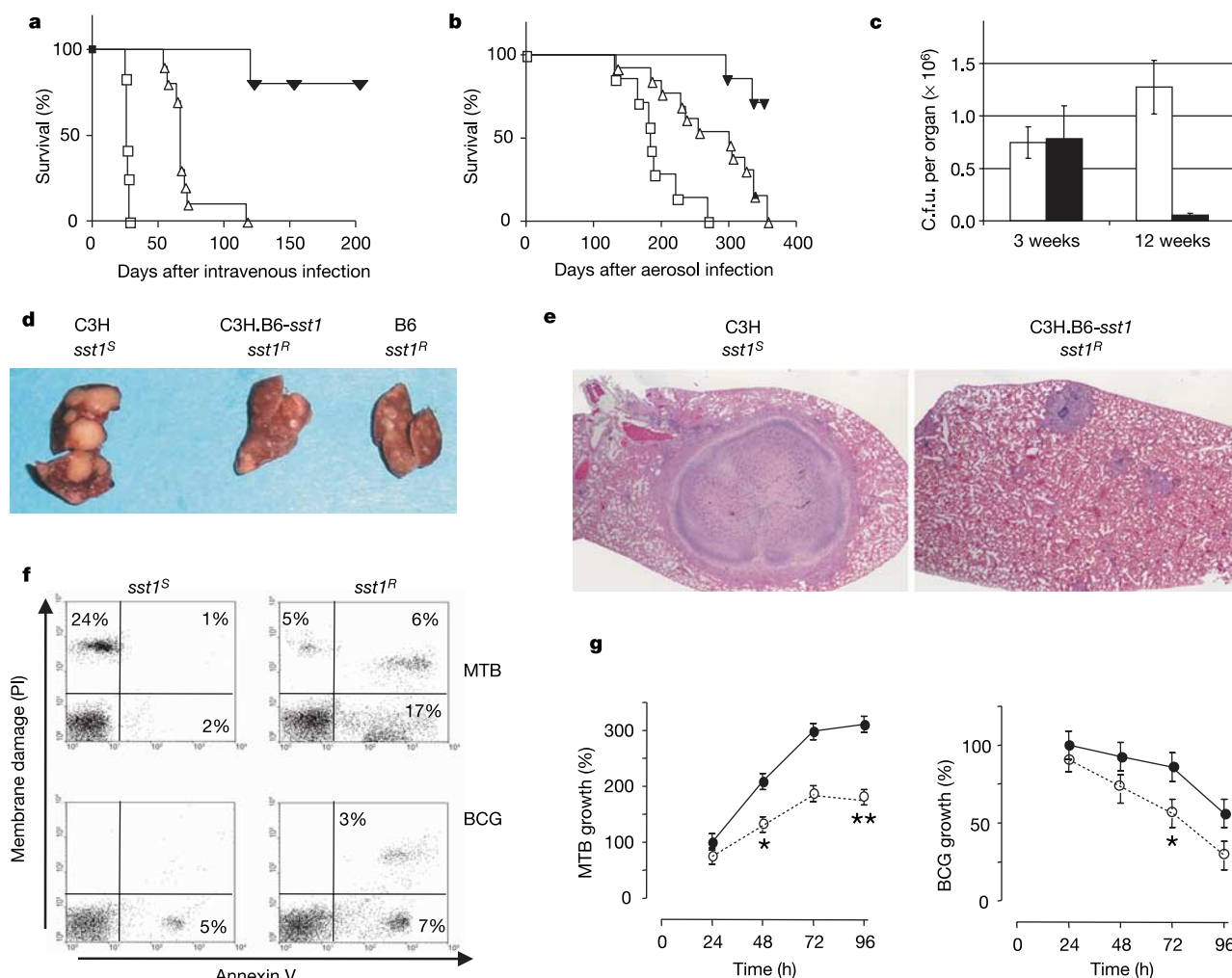


Figure 1 The *sst1* locus mediates innate immunity to tuberculosis. **a**, **b**, Survival of C3H mice (squares), B6 mice (filled triangles) and C3H.B6-*sst1^R* (*sst1^R*) mice (open triangles) after intravenous (**a**) or aerosol (**b**) infection with *M. tuberculosis*. **c**, *M. tuberculosis* bacterial loads in the lungs of the *sst1* congenic mice after aerosol infection. Open bars, *sst1^S*; filled bars, *sst1^R*. **d**, **e**, Tuberculosis lung lesions 25 days after intravenous infection (**d**) and 12 weeks after aerosol infection (**e**). Stain in **e**, haematoxylin/eosin;

original magnification $\times 40$. **f**, FACS analysis of the mechanism of cell death of the *sst1* congenic macrophages infected with *M. tuberculosis* (MTB; top panels) or BCG (bottom panels) *in vitro*. **g**, Multiplication of *M. tuberculosis* (left panel) or *M. bovis* BCG (right panel) in the *sst1* congenic macrophages *in vitro* (asterisk, $P < 0.01$; two asterisks, $P < 0.001$). Open circles, *sst1^S*; filled circles, *sst1^R*. Error bars show 95% confidence intervals.

immediately downstream of the repeat; that is, between the repeat region and the *NppC* gene. A total of 22 known and predicted genes are encoded within the *sst1* critical region according to the Ensembl and Celera databases of the mouse genome (Supplementary Table 1). It was impossible to reduce the *sst1* critical region further by genetic recombination. In the next step we therefore tested the expression of each of the *sst1*-encoded candidate genes in the lungs during tuberculosis infection *in vivo* and in macrophages infected with *M. tuberculosis* *in vitro* using reverse transcriptase-mediated polymerase chain reaction (RT-PCR) and rapid amplification of complementary DNA ends (RACE). From our studies the *Ifi75* gene seemed the most likely candidate (Supplementary Fig. 3b, e).

As shown in Fig. 2b, the 5' RACE products of *Ifi75* in the lungs of tuberculosis-infected mice were strikingly different between the *sst1* congenic strains: a major single band was amplified from the lungs of the *sst1* resistant mice but was absent from the lungs of the *sst1* susceptible strain; instead, multiple weak products were obtained from the latter. Although some aberrant transcripts were present in the lung tissue of the *sst1*^R animals as well, most of the *Ifi75*-related

transcripts in their tuberculosis lung lesions were represented by a single isoform, which we named *Ipr1* to differentiate it from other *Ifi75*-related sequences (*Ifi75*-rs) identified by RACE and perhaps also encoded within the HSR repeat. The predicted *Ipr1* protein is 92% identical to *Ifi75* in *Mus caroli*. It contains an Sp100-like domain in its amino terminus, an LXXLL-type nuclear receptor binding motif (NRB), a bipartite nuclear localization signal (NLS) and a SAND domain (named after Sp100, AIRE-1, NucP41/75 and DEAF-1) at its carboxy terminus (Fig. 2c).

Using DNA probes specific for the Sp100 and SAND domains of the *Ipr1*, we analysed the kinetics of its expression by northern hybridization in the lungs of the *sst1* congenic mouse strains during the progression of tuberculosis (Fig. 2d). Expression of *Ipr1* was detectable in the lungs of the naive *sst1*^R mice, and its expression increased significantly 2 weeks after intravenous infection with *M. tuberculosis* and remained higher at later time points. However, expression of the Sp100- and SAND domain-containing *Ifi75*-rs in the lungs of the *sst1* susceptible C3HeB/FeJ mice remained below the level of detection by northern blot hybridization. Instead, the level of transcripts of another gene encoded within the HSR repeat

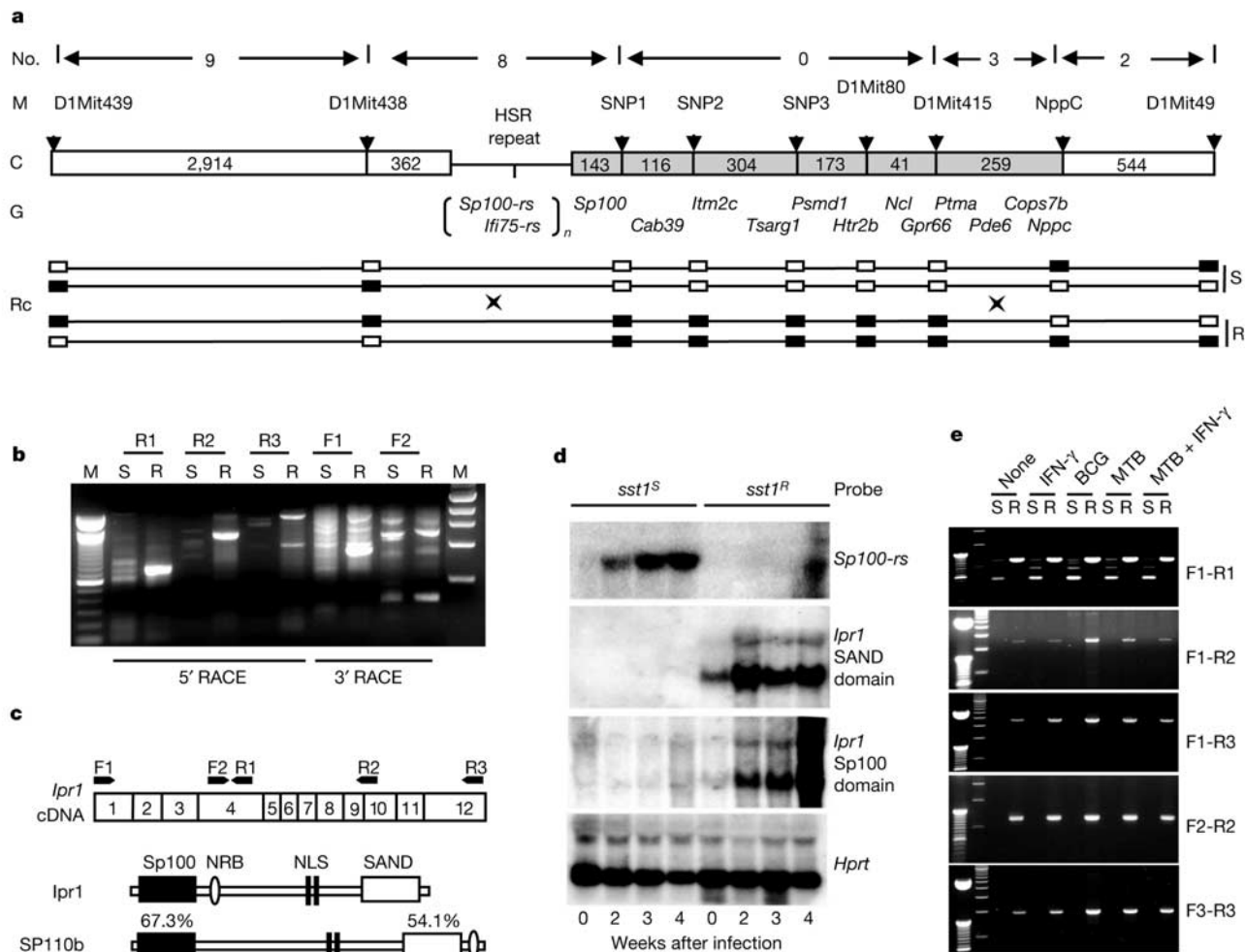


Figure 2 Identification of the *sst1* candidate gene. **a**, Physical map of the *sst1* minimal region. The top line shows the number of recombination events. M, polymorphic markers; C, chromosome with distances between the markers in kilobases; G, known genes; RC, recombinant chromosomes containing the *sst1* resistant (R) or susceptible (S) alleles. Genotypes for each marker are represented by filled (B6) and open (C3H) boxes. **b**, Analysis of the *Ifi75*-rs expression in the tuberculosis lung lesions of the *sst1* congenic

mice by RACE. **c**, Domain structure of *Ipr1* and its human homologue SP110b, and location of the PCR primers. **d**, Expression of *Ipr1* and *Sp100-rs* in the lungs during *M. tuberculosis* infection (northern blot). **e**, Expression of *Ipr1* in *sst1*^S (S) or *sst1*^R (R) macrophages infected with *M. tuberculosis* (MTB) or BCG or activated with interferon- γ (IFN- γ) *in vitro*.

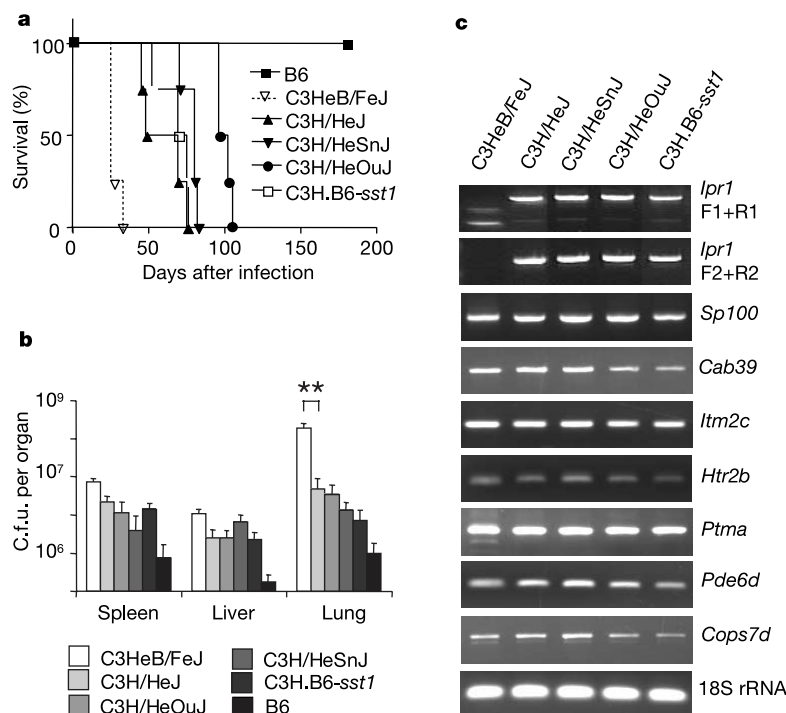


Figure 3 Lack of *Ipr1* expression in the C3HeB/FeJ substrain correlates with its extreme susceptibility to *M. tuberculosis* infection. **a**, Survival after intravenous infection with *M. tuberculosis*. **b**, *M. tuberculosis* bacterial loads 3 weeks after infection (four mice per

strain; two asterisks, $P < 0.001$; error bars show standard deviations). **c**, Analysis by RT-PCR of *sst1*-encoded candidate gene expression in tuberculosis lung lesions 3 weeks after infection.

region, *Sp100*-rs, was elevated in the lungs of the *sst1*^S mice (Fig. 2d).

To investigate the expression of *Ipr1* in macrophages, we used five overlapping combinations of the *Ipr1*-specific PCR primers covering the full-length *Ipr1* transcript (Fig. 2e). *Ipr1* was expressed in non-activated *sst1*^R macrophages, and the level of its expression increased after infection with both avirulent BCG and virulent *M. tuberculosis*. No expression of the full-length transcript of *Ipr1* was seen in the *sst1*^S macrophages under any stimulation conditions. Macrophages isolated from the tuberculosis lung lesions of the *sst1*^R mice also expressed the full-length *Ipr1* transcript, whereas those from the *sst1*^S mice did not (data not shown). We were unable to detect additional *Ipr1*-related transcripts induced in macrophages after infection *in vivo* and *in vitro* or after stimulation with interferon- γ . Even though *Ipr1* is encoded within the HSR repeat region, our data indicate that a single major isoform of this gene might be expressed in *sst1*^R macrophages either before or during tuberculosis infection and that this isoform is not expressed in the *sst1*^S C3HeB/FeJ mice.

The C3HeB/FeJ substrain is unique among all other substrains of C3H mice in terms of its extreme susceptibility to tuberculosis^{10,11}. The C3HeB/FeJ mice die abruptly within 3.5–4 weeks of infection, displaying severe lung pathology. In our experiments, the survival time of other substrains of C3H was considerably longer and was similar to that of the *sst1*^R congenic strain C3H.B6-*sst1* (Fig. 3a). The bacterial loads in the lungs of the C3HeB/FeJ mice at 3 weeks after infection were 50–100-fold higher than in other substrains of C3H and the *sst1*^R congenics (Fig. 3b), possibly indicating a unique allele at the *sst1* locus. We compared the expression of the *sst1*-encoded candidate genes in the lungs of mice of four C3H substrains and the *sst1*^R congenics and found that the lack of expression of *Ipr1* differentiated C3HeB/FeJ from all other substrains of C3H (Fig. 3c). Because all the C3H substrains originated from a common ancestor¹⁷, it is likely that they are genetically identical within the *sst1* region and a unique mutation *de novo* has led to the defect of *Ipr1* expression in C3HeB/FeJ mice and is responsible for a severe defect

in their resistance to tuberculosis.

We generated transgenic mice that expressed a full-length copy of the *Ipr1* cDNA on the susceptible C3HeB/FeJ background in a macrophage-specific manner under the control of the human scavenger receptor A promoter (SR-A). Mature bone marrow-derived macrophages (BMDM), as well as resident peritoneal macrophages obtained from those mice, expressed the *Ipr1* transgene (Fig. 4a). Despite the fact that the regulation of *Ipr1* expression in the transgenic macrophages was clearly less efficient from the SR-A promoter than from the endogenous *Ipr1* promoter (Fig. 4a), when the *Ipr1* transgenic mice were infected with virulent *M. tuberculosis* we observed a statistically significant difference in the bacterial loads between the *sst1*^S (Tg^{-/-}) and the *Ipr1* transgenic (Tg^{+/+}) animals in the lungs (Fig. 4b). *In vitro*, the *Ipr1* transgenic macrophages also controlled the multiplication of *M. tuberculosis* more effectively (Fig. 4c) and turned on the apoptotic pathway of cell death after interaction with virulent *M. tuberculosis* (Fig. 4d, right panels). The growth of another intracellular pathogen, *L. monocytogenes*, was markedly suppressed (50–100-fold) in the *Ipr1* transgenic macrophages (Fig. 4e). As in the *M. tuberculosis* infection, necrotic death accompanied infection of the *sst1*^S macrophages with virulent *L. monocytogenes* (Fig. 4f, left panel), whereas the *Ipr1* transgenic macrophages showed markers of apoptotic death (Fig. 4f, right panel).

Thus the expression of a single gene, *Ipr1*, in the *sst1*^S macrophages restored key functions related to the pathogenesis of tuberculosis that are encoded within the *sst1* locus: greater control of multiplication of virulent *M. tuberculosis* *in vivo* and *in vitro* as well as an apoptotic mechanism of *M. tuberculosis*-induced macrophage cell death. Moreover, *Ipr1* mediates macrophage resistance to another intracellular pathogen, *L. monocytogenes*, indicating that the *Ipr1* product might control a common mechanism of innate resistance against several intracellular pathogens.

The closest homologue of the predicted *Ipr1* protein in humans (41% identity) is SP110b (ref. 18), which localizes to a region of

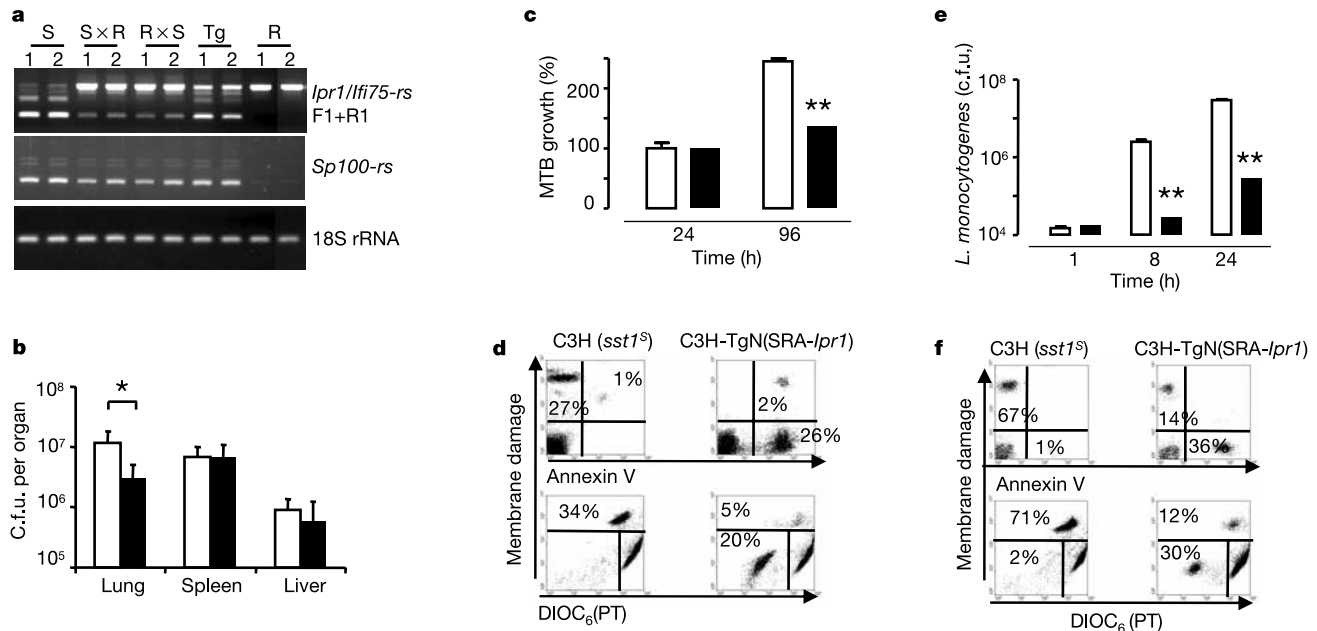


Figure 4 Expression of the *Ipr1* transgene in the *sst1^S* macrophages confers resistance to intracellular pathogens *M. tuberculosis* and *L. monocytogenes*. **a**, RT-PCR of *Ipr1* and *Sp100-rs* in macrophages isolated from *sst1^S* (S) and *sst1^R* (R) mice, their F₁ hybrids (S×R, R×S) and *Ipr1* transgenic (Tg) mice: 1, stimulated with interferon- γ ; 2, infected with *M. tuberculosis*. **b**, Bacterial loads of *M. tuberculosis* in *Ipr1* transgenic (Tg^{+/+}; filled columns) and control (Tg^{-/-}; open columns) mice after infection with *M. tuberculosis* (seven mice per strain; asterisk, $P < 0.05$). **c**, **e**, Growth of *M. tuberculosis* (MTB) (**c**) and *L. monocytogenes* (**e**) in *Ipr1* transgenic (filled columns) and control (*sst1^S*; open

columns) macrophages. Three experiments were performed in triplicate. Asterisk, $P < 0.01$; two asterisks, $P < 0.001$; error bars show 95% confidence intervals. **d**, **f**, FACS analysis of cell death of the *Ipr1* transgenic (right panels) and control (*sst1^S*; left panels) macrophages infected with *M. tuberculosis* (**d**) and *L. monocytogenes* (**f**). Apoptotic cells are Annexin V⁺ (upper panels) and DiOC₆^{low} (lower panels); PT, mitochondrial membrane permeability transition. Necrotic cells are identified by membrane damage (see Methods).

human chromosome 2 syntenic with the *sst1* minimal region on mouse chromosome 1. Both *Ipr1* and the human SP110 proteins contain motifs that are involved in protein–protein interactions (Sp100 domain)^{19,20}, chromatin binding (SAND domain)^{21,22}, nuclear localization signal (NLS) and the NRB motif LXXLL. Recent evidence indicates that human SP110 protein might function as a transcriptional cofactor for nuclear hormone receptors¹⁸ and might directly bind the retinoic acid receptor²³. Signalling through nuclear receptors such as the corticosteroid receptor, retinoic acid receptor, peroxisome-proliferator-activated receptors and vitamin D receptor is important in the control of various aspects of the macrophage life cycle, including differentiation, activation, response to pathogens and apoptosis²⁴. The expression of both *Ipr1* and its human homologue *SP110* is regulated by interferons²⁵, additionally indicating a possible role in immunity in both species. Moreover, polymorphisms in *SP110* have been associated with susceptibility to the hepatitis C virus²⁶, and the SP110b protein has been shown to interact physically with viral proteins such as Epstein–Barr virus SM protein and hepatitis C virus core protein^{23,27}. It is possible that the *Ipr1* and SP110 proteins mediate cross-talk between nuclear receptors, interferon signalling and pathogens. Viruses, and perhaps intracellular pathogens, might have evolved mechanisms to interfere with or exploit the *Ipr1*/SP110 function. Taken together, these data suggest that in mammals, because no *Ipr1* homologues were found in yeasts or insects, the *Ipr1*-related proteins might have a previously undescribed function in integrating signals generated by intracellular pathogens or viruses with mechanisms regulating the activation, gene expression and cell death of host cells²⁸. Therefore, *SP110* might be a candidate gene for testing for association with susceptibility to tuberculosis in human populations. □

Methods

Animals

C57BL/6J, C3HeB/FeJ, C3H/HeJ, C3H/HeOuJ and C3H/HeSnJ inbred mice were obtained from the Jackson Laboratory. The congenic C3H.B6-*sst1* (*sst1^R*), B6.C3H-*sst1* and transgenic C3H-TgN(SRA-*Ipr1*) mouse strains were generated in our laboratory. The C3H.B6-*sst1* congenic mice were obtained by introgression of a roughly 20-centimorgan (cM) segment of B6-derived chromosome 1 with a proximal recombination breakpoint between D1Mit215 (47 cM) and D1Mit334 (49 cM) and a distal limit between D1Mit187 (64 cM) and D1Mit200 (75 cM) on the C3HeB/FeJ genetic background using ten backcrosses. The congenic interval transferred from the B6 resistant background did not include *Slc11a1* (formerly known as *Nramp1*), which is located at 39.2 cM. The *sst1* resistant congenic mouse strain C3H.B6-*sst1* therefore carries the same allele of the *Nramp1* as the parental *sst1* susceptible C3HeB/FeJ mice. The B6.C3H-*sst1* mice were obtained by transferring the *sst1* susceptible allele on the B6 genetic background using ten backcrosses. The transgenic C3H-TgN(SRA-*Ipr1*) mice were established by expressing the C57BL/6J-derived *Ipr1* gene under the control of a macrophage-specific Scavenger Receptor A (SRA) promoter²⁹ on the C3HeB/FeJ genetic background (see Supplementary Methods for details).

Infection of mice with *M. tuberculosis*

For intravenous infection, 10^5 live *M. tuberculosis* were injected into the tail vein in 100 μ l of PBS. Aerosol infections were performed with aerosol apparatus manufactured by the College of Engineering Shops at the University of Wisconsin, Madison. Mice were exposed to aerosol for 20 min, which resulted in the deposition of 15–30 colony-forming units (c.f.u.) per mouse. Mice were killed by halothane anaesthesia. Organs were homogenized in PBS containing 0.05% Tween 80, and serial tenfold dilutions were cultured on 7H10 agar enriched with 10% oleic acid/albumin/dextrose/catalase (OADC; Difco) for 3 weeks at 37 °C.

Isolation and infection of murine BMDM in vitro

BMDM were isolated from femurs and tibias of male C3H, C3H.B6-*sst1* and C3H-TgN(SRA-*Ipr1*) mice (6–8 weeks old) and were infected with *L. monocytogenes* strain 10403S as described previously³⁰. Macrophage monolayers were infected at a multiplicity of infection of 1 *M. tuberculosis* Erdman per 10 macrophages (m.o.i. 1:10). After 6 h the cells were washed with PBS containing 1% FCS (PBS/1% FCS). The cells were incubated in complete medium containing 10% FCS; three coverslips were removed from the culture at indicated time points, and cells on each coverslip were lysed separately with 0.1% Triton X-100. Serial tenfold dilutions of cell lysates were plated on 7H10 agar containing OADC and incubated for 3 weeks at 37 °C.

Differentiation of apoptotic and necrotic pathways

Macrophages were infected at an m.o.i. of 1:10. At indicated time points, cells were stained with 10 nM 3,3'-dihexyloxycarbocyanine iodide (DiOC₆; Molecular Probes) and 0.8 mM ethidium bromide (EB; Sigma) for 20 min at 37°C, washed three times with PBS, fixed for 20 min with 1% paraformaldehyde and washed once more with PBS. Cells were analysed with a BD FACScan flow cytometer (BD Biosciences) to differentiate between live (DiOC₆^{high} EB⁻), apoptotic (DiOC₆^{low} EB⁻) and necrotic (EB⁺) cells. For Annexin V staining, cells were incubated in Annexin binding buffer (10 mM HEPES, 140 mM NaCl, 2.5 mM CaCl₂) and stained for 20 min with 10 µl of Annexin V-Alexa 488 (Molecular Probes), then counterstained with propidium iodide (PI, 1 µg ml⁻¹), washed twice with cold PBS, fixed with 1% paraformaldehyde for 30 min and washed once with PBS. Fluorescence-activated cell sorting (FACS) analysis was used to differentiate between early apoptotic (Annexin V⁺ PI⁻), late apoptotic (Annexin V⁺ PI⁺) and necrotic (Annexin V⁻ PI⁺) cells.

Isolation and analysis of the *Ipr1* cDNA

The *Ipr1*-specific oligonucleotide primers (F1-2 and R1-3) are shown in Supplementary Methods and in Fig. 2c. RACE-PCR was performed with the SMART RACE cDNA Amplification Kit (BD Clontech). The cDNAs were synthesized from the lungs of C3HeB/FeJ and C3H.B6-sst1 mice infected with *M. tuberculosis*. The RACE amplification products were purified with PCR purification columns (Qiagen), cloned into the plasmid vector pGEM-T (Promega) and sequenced with T7 and SP6 primers. A full-length sequence of *Ipr1* was confirmed by sequencing the 'end-to-end' PCR product obtained with the F1 and R3 primers.

Statistical analysis

Statgraphics Plus, release 4, 1999 (Statgraphics Corp.) and GraphPad Prism 3.0 (GraphPad) software were used for the analysis. Comparison of bacterial loads was performed with Student's *t*-test. Results are presented as means ± s.d. The threshold for statistical significance was *P* < 0.05. Kaplan–Meier survival curves were generated and compared by using the log-rank test (GraphPad Prism). Intracellular bacterial growth and cell death were analysed by two-factor analysis of variance (time, genetic backgrounds and experiment). The statistical significance was tested with *P* < 0.05 as the critical value with the Student–Newman–Keuls post-test to compare means between both genetic backgrounds. Data are presented as means ± 95% confidence intervals for the mean.

Received 21 October 2004; accepted 27 January 2005; doi:10.1038/nature03419.

- Raviglione, M. C. The TB epidemic from 1992 to 2002. *Tuberculosis (Edinb.)* **83**, 4–14 (2003).
- Casanova, J. L. & Abel, L. Genetic dissection of immunity to mycobacteria: the human model. *Annu. Rev. Immunol.* **20**, 581–620 (2002).
- Bellamy, R. *et al.* Genetic susceptibility to tuberculosis in Africans: a genome-wide scan. *Proc. Natl Acad. Sci. USA* **97**, 8005–8009 (2000).
- Kramnik, I., Dietrich, W. F., Demant, P. & Bloom, B. R. Genetic control of resistance to experimental infection with virulent *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **97**, 8560–8565 (2000).
- Bleed, D., Dye, C. & Raviglione, M. C. Dynamics and control of the global tuberculosis epidemic. *Curr. Opin. Pulm. Med.* **6**, 174–179 (2000).
- Corbett, E. L. *et al.* The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. *Arch. Intern. Med.* **163**, 1009–1021 (2003).
- Bloom, B. R. Tuberculosis—the global view. *N. Engl. J. Med.* **346**, 1434–1435 (2002).
- Cosma, C. L., Sherman, D. R. & Ramakrishnan, L. The secret lives of the pathogenic mycobacteria. *Annu. Rev. Microbiol.* **57**, 641–676 (2003).
- Taylor, J. L. *et al.* Pulmonary necrosis resulting from DNA vaccination against tuberculosis. *Infect. Immun.* **71**, 2192–2198 (2003).
- Kamath, A. B., Alt, J., Debbabi, H. & Behar, S. M. Toll-like receptor 4-defective C3H/HeJ mice are not more susceptible than other C3H substrains to infection with *Mycobacterium tuberculosis*. *Infect. Immun.* **71**, 4112–4118 (2003).
- Kramnik, I., Demant, P. & Bloom, B. B. Susceptibility to tuberculosis as a complex genetic trait: analysis using recombinant congenic strains of mice. *Novartis Found. Symp.* **217**, 120–131 (1998).
- Hsu, T. *et al.* The primary mechanism of attenuation of bacillus Calmette–Guérin is a loss of secreted lytic function required for invasion of lung interstitial tissue. *Proc. Natl Acad. Sci. USA* **100**, 12420–12425 (2003).
- Guinn, K. M. *et al.* Individual RD1-region genes are required for export of ESAT-6/CFP-10 and for virulence of *Mycobacterium tuberculosis*. *Mol. Microbiol.* **51**, 359–370 (2004).
- Agulnik, S., Plass, C., Traut, W. & Winking, H. Evolution of a long-range repeat family in chromosome 1 of the genus *Mus*. *Mamm. Genome* **4**, 704–710 (1993).
- Traut, W., Rahn, I. M., Winking, H., Kunze, B. & Weichenhan, D. Evolution of a 6–200 Mb long-range repeat cluster in the genus *Mus*. *Chromosoma* **110**, 247–252 (2001).
- Weichenhan, D. *et al.* Source and component genes of a 6–200 Mb gene cluster in the house mouse. *Mamm. Genome* **12**, 590–594 (2001).
- Krog, H. H. & Moutier, R. Identification of inbred strains of mice. II. Characterization of different substrains of the C3H strain. *J. Hered.* **69**, 66–70 (1978).
- Bloch, D. B. *et al.* Sp110 localizes to the PML–Sp100 nuclear body and may function as a nuclear hormone receptor transcriptional coactivator. *Mol. Cell. Biol.* **20**, 6138–6146 (2000).
- Wasyluk, C., Schlumberger, S. E., Ciqui-Filipe, P. & Wasyluk, B. Sp100 interacts with ETS-1 and stimulates its transcriptional activity. *Mol. Cell. Biol.* **22**, 2687–2702 (2002).
- Sternsdorf, T., Jensen, K., Reich, B. & Will, H. The nuclear dot protein sp100, characterization of domains necessary for dimerization, subcellular localization, and modification by small ubiquitin-like modifiers. *J. Biol. Chem.* **274**, 12555–12566 (1999).
- Surdo, P. L., Bottomley, M. J., Sattler, M. & Scheffzek, K. Crystal structure and nuclear magnetic resonance analyses of the SAND domain from glucocorticoid modulatory element binding protein-1 reveals deoxyribonucleic acid and zinc binding regions. *Mol. Endocrinol.* **17**, 1283–1295 (2003).
- Bottomley, M. J. *et al.* The SAND domain structure defines a novel DNA-binding fold in transcriptional regulation. *Nature Struct. Biol.* **8**, 626–633 (2001).

- Watahi, K. *et al.* Modulation of retinoid signaling by a cytoplasmic viral protein via sequestration of Sp110b, a potent transcriptional corepressor of retinoic acid receptor, from the nucleus. *Mol. Cell. Biol.* **23**, 7498–7509 (2003).
- Castrillo, A. & Tontonoz, P. Nuclear receptors in macrophage biology: at the crossroads of lipid metabolism and inflammation. *Annu. Rev. Cell. Dev. Biol.* **20**, 455–480 (2004).
- Kadereit, S., Gewert, D. R., Galabru, J., Hovanessian, A. G. & Meurs, E. F. Molecular cloning of two new interferon-induced, highly related nuclear phosphoproteins. *J. Biol. Chem.* **268**, 24432–24441 (1993).
- Saito, T. *et al.* Genetic variations in humans associated with differences in the course of hepatitis C. *Biochem. Biophys. Res. Commun.* **317**, 335–341 (2004).
- Nicewonger, J., Suck, G., Bloch, D. & Swaminathan, S. Epstein–Barr virus (EBV) SM protein induces and recruits cellular Sp110b to stabilize mRNAs and enhance EBV lytic gene expression. *J. Virol.* **78**, 9412–9422 (2004).
- Hofmann, T. G. & Will, H. Body language: the function of PML nuclear bodies in apoptosis regulation. *Cell Death Differ.* **10**, 1290–1299 (2003).
- Horvai, A. *et al.* Scavenger receptor A gene regulatory elements target gene expression to macrophages and to foam cells of atherosclerotic lesions. *Proc. Natl Acad. Sci. USA* **92**, 5391–5395 (1995).
- Boyartchuk, V. *et al.* The host resistance locus *ss1* controls innate immunity to *Listeria monocytogenes* infection in immunodeficient mice. *J. Immunol.* **173**, 5112–5120 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank I. Breiterene, K. Vasquez, K. Sigrist and C. Mottley for technical assistance, and R. Kucherlapati, P. Demant, W. F. Dietrich, S. Agoulnik and D. Bloch for discussions and support throughout the project. This work was supported by the National Institutes of Health.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to I.K. (ikramnik@hsph.harvard.edu). The *Ipr1* sequence has been deposited in GenBank under accession number AY845948.

IRF-7 is the master regulator of type-I interferon-dependent immune responses

Kenya Honda^{1*}, Hideyuki Yanai^{1*}, Hideo Negishi¹, Masataka Asagiri¹, Mitsuharu Sato², Tatsuaki Mizutani¹, Naoya Shimada¹, Yusuke Ohba^{1,3}, Akinori Takaoka¹, Nobuaki Yoshida² & Tadatsugu Taniguchi¹

¹Department of Immunology, Graduate School of Medicine and Faculty of Medicine, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan

²Institute of Medical Sciences, University of Tokyo, Shirokanedai 4-6-1, Minato-ku, Tokyo 108-8639, Japan

³Information and Cell Function, PRESTO, JST, Kawaguchi, Saitama 332-0012, Japan

* These authors contributed equally to this work

The type-I interferon (IFN-α/β) response is critical to immunity against viruses and can be triggered in many cell types by cytosolic detection of viral infection, or in differentiated plasmacytoid dendritic cells by the Toll-like receptor 9 (TLR9) subfamily, which generates signals via the adaptor MyD88 to elicit robust IFN induction^{1–4}. Using mice deficient in the *Irf7* gene (*Irf7*^{-/-} mice), we show that the transcription factor IRF-7 is essential for the induction of IFN-α/β genes via the virus-activated, MyD88-independent pathway and the TLR-activated, MyD88-dependent pathway. Viral induction of MyD88-independent IFN-α/β genes is severely impaired in *Irf7*^{-/-} fibroblasts. Consistently, *Irf7*^{-/-} mice are more vulnerable than *Myd88*^{-/-} mice to viral infection, and this correlates with a marked decrease in serum IFN levels, indicating the importance of the IRF-7-dependent induction of systemic IFN responses for innate antiviral immunity. Furthermore, robust induction of IFN production by activation of the TLR9 subfamily in plasmacytoid dendritic cells is entirely dependent on IRF-7, and this MyD88–

IRF-7 pathway governs the induction of CD8⁺ T-cell responses. Thus, all elements of IFN responses, whether the systemic production of IFN in innate immunity or the local action of IFN from plasmacytoid dendritic cells in adaptive immunity, are under the control of IRF-7.

The induction of IFN- α/β production is primarily controlled at the transcriptional level, wherein IRF-7 of the IRF family of transcription factors has been the focus of attention^{1,5–7}. However,

the actual contribution of IRF-7 to the transcriptional activation of IFN- α/β genes in distinct cell types upon viral infection or TLR activation is still not clarified. Furthermore, the *in vivo* roles of IRF-7 in the regulation of innate and adaptive immunity remain unknown.

We generated *Irf7*^{−/−} mice by the standard homologous recombination protocol, and their nullizygosity was confirmed by DNA and RNA blot and immunoblot analyses (Fig. 1a). The mutant mice

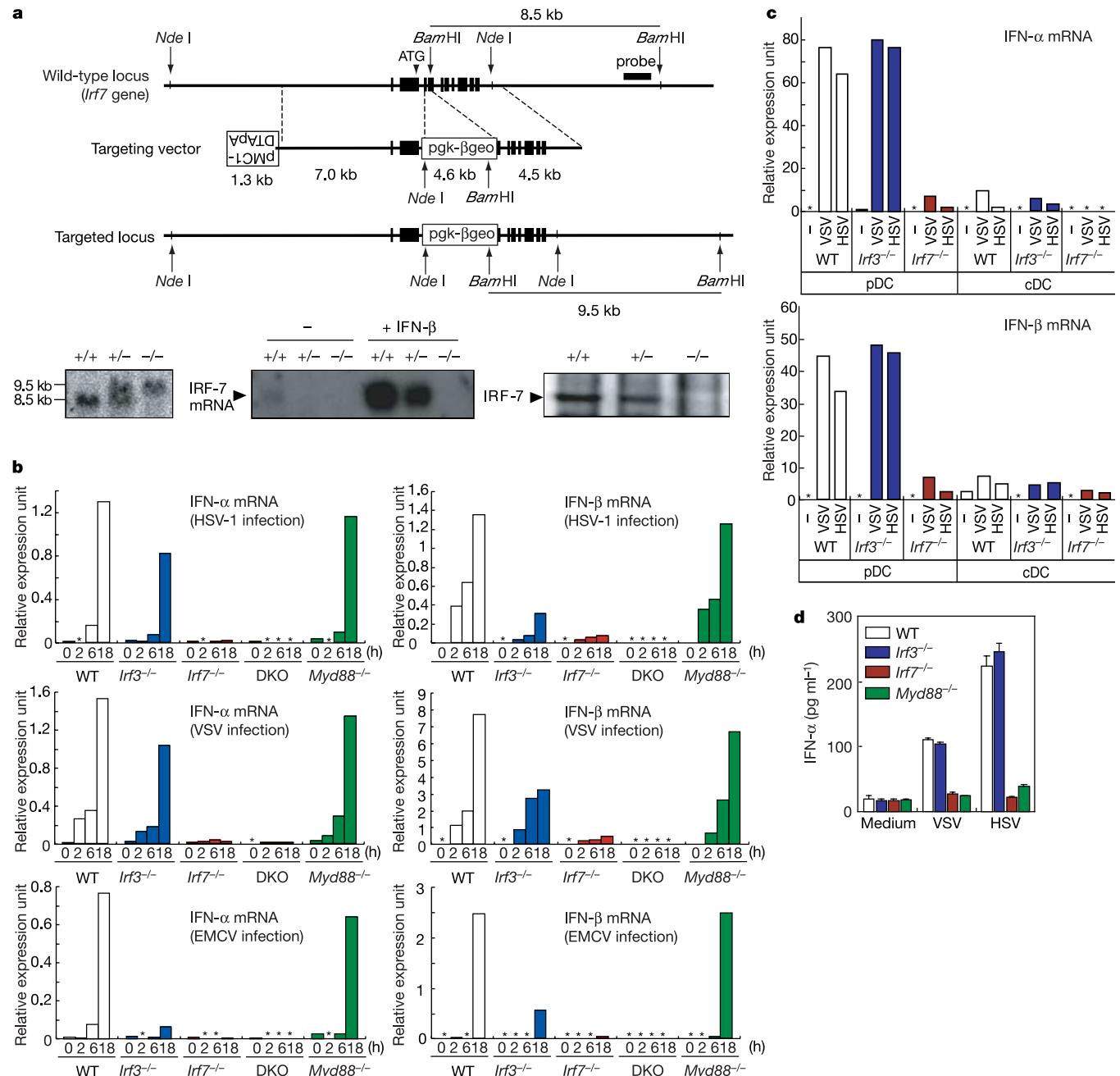


Figure 1 Impaired induction of IFN- α/β gene expression in *Irf7*^{−/−} MEFs and pDCs by viral infection. **a**, Generation of mice with *Irf7* nullizygosity. Partial restriction map of the wild-type locus and targeting strategy for *Irf7* gene disruption are displayed. The probe for genomic Southern blot analysis is indicated by a thick bar. The bottom left panel shows a Southern blot analysis of *Bam*HI-digested genomic DNA isolated from wild-type (+/+), heterozygous (+/-) and homozygous mutant (-/-) mice. The results of RNA blotting and immunoblotting analyses of MEFs treated with IFN- β (500 U ml⁻¹) are shown in bottom middle and right panels, respectively. **b**, MEFs from various mutant mice were infected with the respective viruses for the indicated periods. IFN- α or IFN- β mRNA levels

were determined by quantitative real-time RT-PCR analysis. Asterisk, not detected. Until 18 h, there was no significant difference in viral replication between these mutant cells (see Supplementary Fig. 2d). DKO, *Irf7*^{−/−}/*Irf3*^{−/−} double knockout. **c**, Wild-type, *Irf3*^{−/−} and *Irf7*^{−/−} mice were infected by HSV-1 or VSV. Nine hours after infection, splenic B220⁺/CD11c^{int} pDCs and B220⁺/CD11c⁺ cDCs were purified by FACS, and IFN- α/β mRNA levels were determined by quantitative real-time RT-PCR analysis. **d**, B220⁺/CD11c^{int} pDCs were purified from wild-type or mutant splenocytes and infected with HSV-1 or VSV for 24 h. IFN- α concentration was measured by ELISA. Results shown are the means ($\pm$ s.d.) of triplicate determinations.

developed normally and no overt differences were observed in haematopoietic cell populations (Supplementary Fig. 1a–c). We first examined IFN- α/β induction in mouse embryonic fibroblasts (MEFs) infected by one of three distinct viruses: herpes simplex virus type 1 (HSV-1), encephalomyocarditis virus (EMCV) and vesicular stomatitis virus (VSV). As shown in Fig. 1b, IFN- α messenger RNA induction was abolished in *Irf7*^{-/-} MEFs, as revealed by quantitative polymerase chain reaction with reverse transcription (RT-PCR) (see also Supplementary Fig. 2a). Similarly, the induction of IFN- β mRNA was markedly inhibited in *Irf7*^{-/-} MEFs, and the residual mRNA induction was abrogated in MEFs from mice with an additional deficiency for the *Irf3* gene (*Irf7*^{-/-}/*Irf3*^{-/-} MEFs or double knockout MEFs⁷, Fig. 1b; see also Supplementary Fig. 2b). Consistently, total IFN activity in the supernatant of *Irf7*^{-/-} MEFs was markedly reduced (Supplementary Fig. 2c). In contrast, MEFs lacking *Myd88* retained the ability to induce IFN- α/β mRNAs in response to these three viruses (Fig. 1b).

These results indicate the operation of the MyD88-independent but IRF-7-dependent pathway for IFN- α/β gene induction. This pathway, which we refer to hereafter as the ‘classical pathway’, has been extensively studied in the context of innate antiviral response, wherein molecules such as double-stranded RNA-activated protein kinase (PKR⁸), retinoic-acid-inducible gene-I (RIG-I; ref. 9), IkB kinase ϵ and TANK-binding kinase 1 (IKK- ϵ and TBK1, respectively; refs 10, 11), and the adaptor Fas-associated death domain (FADD¹²) are involved. Although IRF-3 also participates in this classical pathway (refs 7, 9–12; see also Fig. 1b and Supplementary Fig. 2c), it contributes little in the absence of IRF-7, presumably because IRF-3 needs to interact with IRF-7 for its full function.

Plasmacytoid dendritic cells (pDCs; also referred to as IFN-producing dendritic cells) stand out as high producers of IFN- α/β after activation of TLR9 and its subfamily member TLR7(8)^{3,4,13–17}. Although there is much circumstantial evidence for the potential role of IRF-7 in the MyD88-dependent induction of IFN- α/β genes

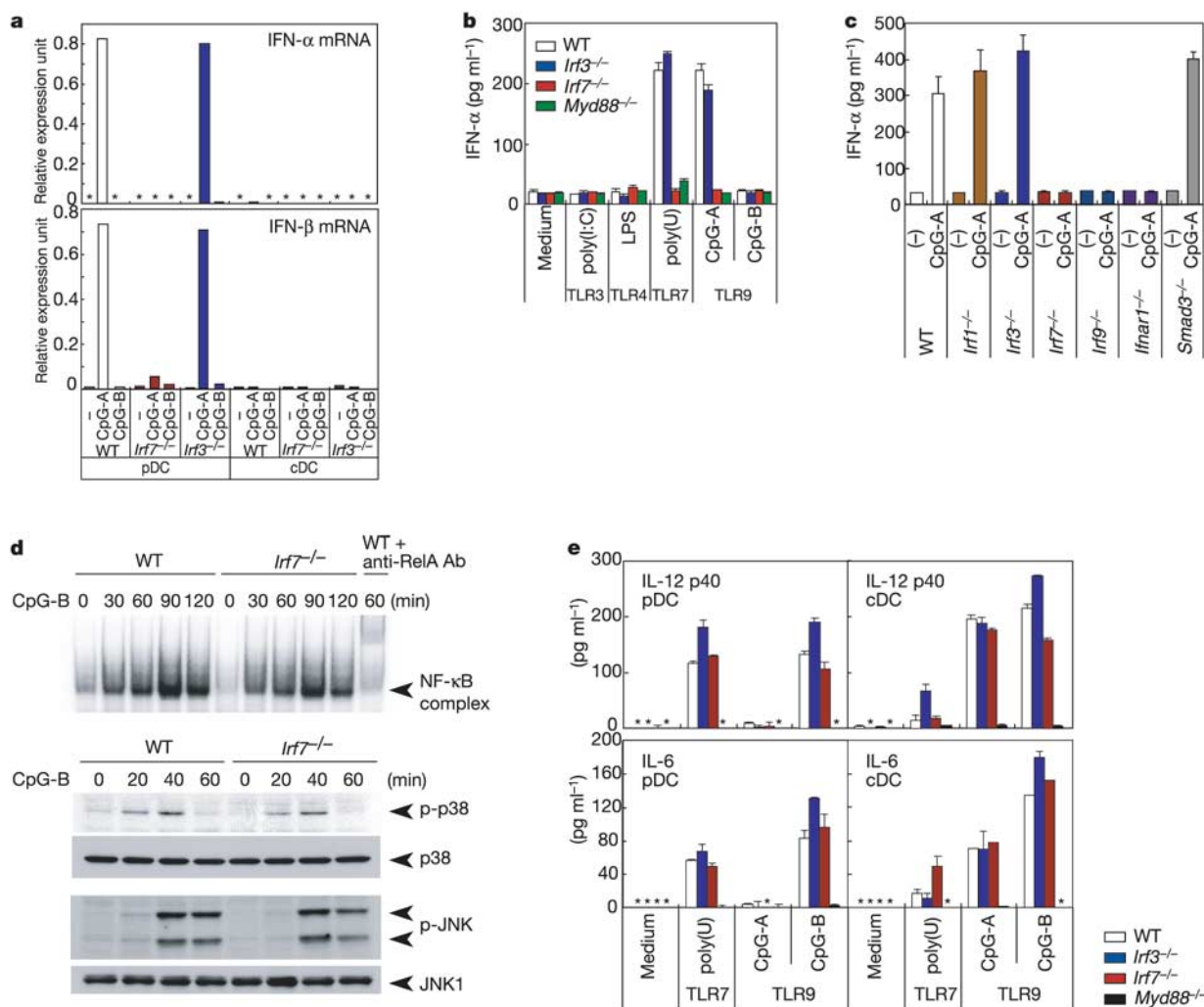


Figure 2 The role of IRF-7 in MyD88-dependent signalling. **a**, Wild-type, *Irf3*^{-/-} and *Irf7*^{-/-} splenic pDCs and cDCs were stimulated with CpG-A or CpG-B for 12 h. IFN- α (top panel) or IFN- β (bottom panel) mRNA levels were analysed by real-time RT-PCR analysis. Asterisk, not detected. **b**, Splenic pDCs were stimulated with the indicated stimuli, and 24 h later IFN- α concentration was measured by ELISA. **c**, Splenic pDCs from wild-type and various mutant mice (*Irf1*^{-/-}, *Irf3*^{-/-}, *Irf7*^{-/-}, *Irf9*^{-/-}, *Irfar1*^{-/-}, *Myd88*^{-/-} and *Smad3*^{-/-} mice) were stimulated with CpG-A for 24 h. IFN- α

concentration was measured by ELISA. **d**, Flt3L-cultured bone marrow-derived dendritic cells from wild-type or *Irf7*^{-/-} mice were stimulated with CpG-B. At the indicated time points, the cells were lysed; NF- κ B activation was analysed by EMSA (top panel) and the phosphorylation of p38 and JNK was analysed by western blotting (bottom panel). **e**, Wild-type or mutant splenic dendritic cells were stimulated with poly(U) plus DOTAP, CpG-A and CpG-B for 24 h. IL-6 and IL-12 p40 concentrations were measured by ELISA. Results shown are the means ($\pm$ s.d.) of triplicate determinations for **b**, **c** and **e**.

as a result of TLR activation in pDCs^{18,19}, it is still unclear to what extent IRF-7 participates in the robust IFN production mechanism characteristic for this cell type. We then purified splenic pDCs from wild-type and mutant mice by cell sorting (>95% purity; see also Supplementary Fig. 3a), and infected them with HSV-1 and VSV, which activate TLR9 and TLR7(8) (the TLR9 subfamily), respectively^{13,16}. As shown in Fig. 1c, the induction of IFN- α and - β mRNA upon viral infection was impaired in *Irf7*^{-/-} pDCs, but was normal in *Irf3*^{-/-} pDCs. Consistently, robust IFN- α production is abolished in *Irf7*^{-/-} pDCs but not in *Irf3*^{-/-} pDCs (Fig. 1d; see also Supplementary Fig. 3b). As reported previously^{13,16}, the production of IFN- α as a result of infection by these viruses is completely dependent on MyD88 in splenic pDCs (Fig. 1d; see also Supplementary Fig. 3b, c). These results indicate that the pathway for robust IFN production in pDCs is distinct from the classical pathway and is subject to the MyD88-dependent induction of IFN- α/β gene transcription, which is also mediated by IRF-7.

We next examined the induction of IFN- α/β mRNA expression in spleen-derived dendritic cell subsets by stimulation with CpG-A, an oligodeoxynucleotide containing the unmethylated CpG motif (D19; ref. 20) that activates TLR9, resulting in robust IFN- α production¹⁷. As shown in Fig. 2a, robust induction of IFN- α/β mRNA expression was observed upon CpG-A stimulation of pDCs, but not conventional DCs (cDCs), from wild-type mice, and this induction was abolished in pDCs from *Irf7*^{-/-} mice (Fig. 2a; see also Supplementary Fig. 3d) despite the normal expression of TLR9 mRNA (Supplementary Fig. 3e). In contrast, induction of IFN- α/β mRNA expression occurred normally in *Irf3*^{-/-} pDCs (Fig. 2a). As shown in Fig. 2b, the CpG-A-induced production of IFN- α was abolished in *Irf7*^{-/-} and *Myd88*^{-/-} pDCs, whereas it was normal in *Irf3*^{-/-} pDCs (see also Supplementary Fig. 3f). The robust production of IFN- α observed upon stimulation with the synthetic RNA polyuridylic acid (poly(U)), which activates the TLR9 subfamily TLR7(8)^{14,15}, was also abolished in *Irf7*^{-/-} and *Myd88*^{-/-} pDCs, but was normal in *Irf3*^{-/-} pDCs (Fig. 2b). Taken together, these results indicate that IRF-7 is essential and IRF-3 is dispensable

for the MyD88-dependent induction of IFN- α/β genes via the TLR9 subfamily. Essentially the same results were obtained with *in vitro* cultured bone-marrow-derived pDCs (data not shown). Other transcription factors, such as IRF-1, IRF-5 and Smad3, are also implicated in IFN gene induction^{1,21} but they seem to be dispensable, because IFN- α production induced by CpG-A was normal in pDCs from mice lacking either of these factors (Fig. 2c; see also ref. 22).

It is worth noting that robust production of IFN- α was abolished in *Ifnar1*^{-/-} and *Irf9*^{-/-} pDCs (Fig. 2c), both of which are defective in the IFN-signal-dependent induction of *Irf7*⁵⁻⁷. Thus, similar to the induction of IFN genes by viruses in MEFs⁵⁻⁷, a positive feedback mechanism also constitutes an essential aspect in pDCs; that is, activation of basally expressed IRF-7 initially induces IFN production at a low level, which subsequently enhances IFN-signal-dependent IRF-7 expression to sustain IFN gene transcription. In fact, IRF-7 mRNA is expressed in unstimulated pDCs and other cell types (Supplementary Fig. 3g). Although the IFN signal is dispensable for an early phase of IFN- α mRNA induction in pDCs (ref. 23; see also Supplementary Fig. 3h), prolonged activation of the TLR9–MyD88 signalling pathway and IFN-dependent IRF-7 induction are required for the robust production of IFN- α .

Because TLR9-dependent activation of NF- κ B, stress-activated protein kinase (SAPK)/JNK and p38 mitogen-activated protein kinase (MAP) kinase is dependent on the canonical MyD88–TRAF6 pathway² and IRF-7 interacts with both MyD88 and TRAF6 (refs 18, 19), we asked whether IRF-7 is also required for the activation of these pathways. As shown in Fig. 2d, activation of NF- κ B, JNK and p38 occurs normally in *Irf7*^{-/-} dendritic cells when stimulated by CpG-B (oligodeoxynucleotide 1668; ref. 24). Consistently, induction of pro-inflammatory cytokines such as interleukin-12 (IL-12) and IL-6 is not inhibited in *Irf3*^{-/-} and *Irf7*^{-/-} dendritic cells (Fig. 2e). We previously proposed the presence of a cytoplasmic transductional-transcriptional processor (CTTP), in which MyD88 forms a multi-molecular complex with IRF-7, TRAF6 and IRAK4 (ref. 18). More recently, we reported that another IRF member, IRF-5, also interacts with and is activated by MyD88, and that IRF-5 is essential for pro-inflammatory cytokine induction²². Together, these results suggest that the function of CTTP in the activation of the NF- κ B/MAP kinase and IRF-5 pathways is independent of IRF-7, which selectively regulates the IFN limb of the MyD88-dependent cytokine gene induction programme in TLR signalling. As such, the TLR9/7–MyD88–IRF-7 pathway for IFN- α/β gene induction genetically defined here is distinct from IFN- β gene induction through the activation of other TLRs, such as TLR4, wherein activation of IRF-3 by TRIF (Toll/IL-1 receptor domain-containing adaptor inducing IFN- β) and TRAM (TRIF-related adaptor molecule) has a critical part (ref. 2; see also Supplementary Fig. 4a).

We next examined the role of IRF-7 in the *in vivo* IFN response against infections by DNA and RNA viruses. As shown in Fig. 3a, c, IFN- α induction is markedly inhibited in the sera of *Irf7*^{-/-} mice infected with either of these viral types, whereas it remained high in both *Irf3*^{-/-} and *Myd88*^{-/-} mice, highlighting the major contribution of the classical MyD88-independent, IRF-7-dependent pathway to systemic IFN induction against viral infection. Consistently, *Irf7*^{-/-} mice are more vulnerable than *Irf3*^{-/-} and *Myd88*^{-/-} mice to these viral infections (Fig. 3b, d). It has been shown that although IFN production is dependent on the TLR9–MyD88 pathway in HSV-1-infected splenic pDCs, this pathway is not essential for the *in vivo* innate immune response against this virus^{13,25}. These previous reports and our current observations suggest that the classical IFN- α/β induction pathway operational in many cell types (see Fig. 1b and Supplementary Fig. 3c)—which is less effective compared with the MyD88–IRF-7-dependent pathway activated by the TLR9 subfamily in pDCs—constitutes a critical part of the innate antiviral defence, wherein IRF-7 has the essential role.

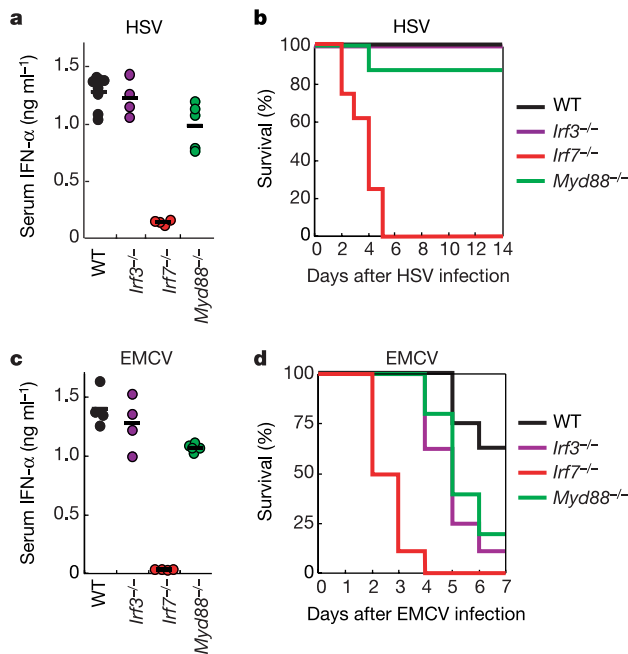


Figure 3 Role of IRF-7 in the *in vivo* IFN response against viral infections. **a–d**, Wild-type and the indicated mutant mice ($n = 8$ per group) were infected with HSV-1 or EMCV. Sera from each mouse were collected at 12 h after infection and IFN- α levels were measured by ELISA (**a**, **c**). The bars indicate the mean values. The survival of these mice was monitored for 14 days (**b**, for HSV-1) and 7 days (**d**, for EMCV).

Because TLR9 signalling is the focus of much attention in the context of the regulation of adaptive immune responses^{26,27}, we next examined the role of the MyD88–IRF-7 pathway in the induction of the antigen-specific CD8⁺ T-cell response by immunizing mice with soluble ovalbumin and TLR agonists²⁸. As shown in Fig. 4a, induction was severely impaired in both *Irf7*^{−/−} and *Myd88*^{−/−} mice when CpG-A was used as the adjuvant. Furthermore, the CpG-A-dependent, ovalbumin-specific CD8⁺ T-cell response was also impaired by pre-treatment of wild-type mice with 120G8, an antibody that allows selective depletion of pDCs (ref. 29; see also Fig. 4b). On the other hand, CpG-A-dependent antibody responses against 2,4,6-trinitrophenol-keyhole limpet haemocyanin remained unaffected in *Irf7*^{−/−} mice (Supplementary Fig. 4b), which is congruent with the above results showing the normal induction of pro-inflammatory cytokines in *Irf7*^{−/−} dendritic cells. These results collectively demonstrate the selective and essential role of the MyD88–IRF-7 pathway in pDCs in the TLR9-mediated triggering of the CD8⁺ T-cell response, which is mediated by IFN production^{28,30}. It is worth noting that the CD8⁺ T-cell response was normal in *Irf7*^{−/−} mice when the adjuvant used was a mycoplasma lipopeptide (MALP2) that activates TLR2 (and TLR6) but does not induce high-level IFN production, indicating the operation of the

MyD88-dependent, IRF-7-independent gene activation programme for this TLR-induced T-cell response (ref. 28; see also Fig. 4a).

Our study points to the differential contribution of the MyD88-independent systemic IFN response by the classical pathway compared with the MyD88-dependent local IFN response by the TLR pathway for innate antiviral immunity and CD8⁺ T-cell adaptive immunity, respectively. However, it must be emphasized that this functional dichotomy may not be applicable to every case of the immune response. As such, depending on the nature of the virus or viral load, both pathways may need to cooperate to ensure a robust antiviral response. In any case, our results clearly demonstrate that IRF-7 governs the entire IFN response. The detailed mechanism of IRF-7 activation by TLR9 signalling remains to be clarified, although our previous study suggests the involvement of the IRAK kinases for IRF-7 phosphorylation¹⁸. Perhaps more interestingly, the mechanism of how the robust IFN production is achieved in pDCs but not in cDCs in response to the same TLR stimuli remains enigmatic, and this is an interesting future issue to be addressed. □

Methods

Generation of *Irf7*^{−/−} mice

Genomic DNA containing the *Irf7* gene was isolated from a 129/Sv mouse genomic library. An *Irf7* gene-targeting construct that replaces exons 2 and 3 (corresponding to amino acids 7–109; 383 base pairs) with a phosphoglycerate kinase promoter-driven β -geo-positive selection cassette (*pgk*– β -geo) was transfected into E14-1 ES cells. We microinjected two independent homologous recombinants into C57BL/6 blastocysts and intercrossed heterozygous F₁ progenies to obtain *Irf7*^{−/−} mice. Mice from these independent clones displayed identical phenotypes.

Other mutant mice

The generation of *Irf1*^{−/−}, *Irf3*^{−/−} and *Irf9*^{−/−} mice has been described previously¹⁷. *Ifnar1*^{−/−} mice were purchased from B&K Universal Group. *Smad3*^{−/−} and *Myd88*^{−/−} mice were provided by K. Yokote and S. Akira, respectively. All the mice were maintained under specific pathogen-free conditions in the animal facility of the University of Tokyo and used after backcrossing with C57BL/6 mice eight to ten times.

Reagents

Synthesized oligodeoxynucleotides were purchased from Hokkaido System Science. The sequences of oligodeoxynucleotides are as follows: CpG-A (D19; ref. 20), ggTGCATCGATGCAgggggG; CpG-B (oligodeoxynucleotide 1668; ref. 24), tccatgacgttcctgatgct. Upper-case and lower-case letters indicate bases with phosphodiester- and phosphorothioate-modified backbones, respectively. The pDC-specific monoclonal antibody (120G8; ref. 29) and IFN- β were provided by G. Trinchieri and Toray industries, respectively.

Preparation of dendritic cells

Spleens were digested with 1 mg ml^{−1} collagenase A (Roche Biochemicals) and 20 mM EDTA, and subjected to negative selection of T and B cells with anti-CD5 and anti-CD19 antibodies (eBioscience), and anti-rat IgG-coated Dynabeads (Dyna). Recovered cells were incubated with anti-B220 and anti-CD11c antibodies (BD Biosciences). B220⁺/CD11c⁺ cDCs and B220⁺/CD11c^{int} pDCs (where superscript 'int' indicates intermediate) were sorted using FACS Diva (BD Bioscience). To prepare bone-marrow-derived dendritic cells, bone marrow cells were cultured with 100 ng ml^{−1} human Flt3L (PeproTech) for 6 days.

RNA analysis

Total RNA was prepared as described previously⁷. Quantitative real-time RT-PCR analysis was performed using LightCycler and SYBRGreen system (Roche). Data were normalized by the level of β -actin expression in each individual sample. Primers for β -actin and IFN- β have been described previously⁷. Primers for IFN- α 1 were used as follows: 5'-GCCTTGACACTCCTGGTACAAATGAG-3' (sense) and 5'-CAGCACATTGGCAGAGGAAGACAG-3' (anti-sense).

Measurement of cytokine production

Cells were seeded onto 96-well plates at 2×10^5 cells ml^{−1} and stimulated for 24 h with various reagents as follows: poly(U) (5 μ g ml^{−1}, Sigma); LPS from *Salmonella minnesota* Re-595 (100 ng ml^{−1}, Sigma); poly(I:C) (100 μ g ml^{−1}, Amersham); CpG-A (3 μ M); and CpG-B (3 μ M). Poly(U) was complexed with DOTAP (Roche Diagnostics) according to the manufacturer's instruction. Cytokine concentration in the supernatants was measured using an enzyme-linked immunosorbent assay (ELISA). The ELISA kit for mouse IFN- α was purchased from PBL Biomedical Laboratories. ELISA kits for mouse IL-12 p40 and IL-6 were obtained from TECHNE Corp.

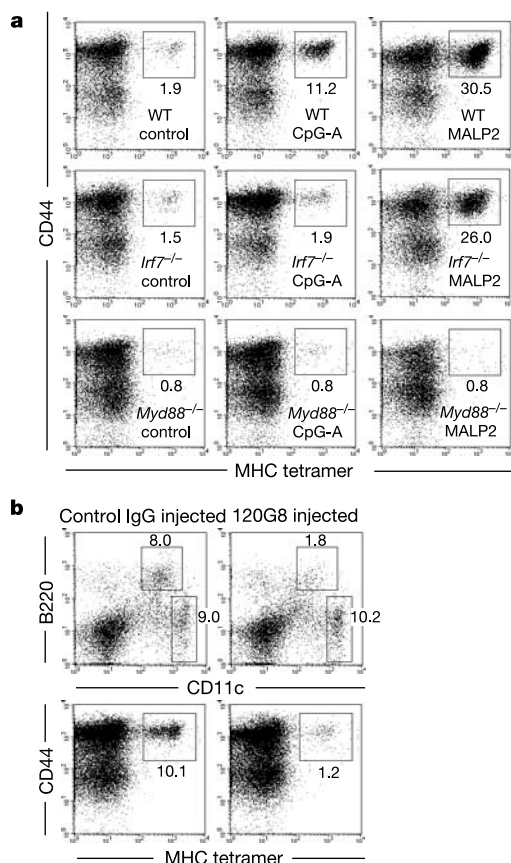


Figure 4 Immune responses of *Irf7*^{−/−} mice. **a**, Wild-type, *Irf7*^{−/−} or *Myd88*^{−/−} mice were immunized with soluble ovalbumin and TLR agonist (CpG-A or MALP2) combined with a CD40 agonistic antibody in order to synergistically stimulate the CD8⁺ T-cell response²⁸. Six days later, splenic cells were isolated and subjected to three-colour FACS analysis using anti-CD8 α antibody, anti-CD44 antibody and the MHC tetramer. The data shown were gated on CD8 α -positive events. The numbers indicate the percentage of tetramer-positive cells relative to the total number of CD8 α ⁺ T cells. Data are representative of three experiments. **b**, Spleen cells from mice treated with pDC-specific antibody (120G8) or control rat IgG were collected 24 h after antibody treatment and analysed for dendritic cell populations by FACS (top panel). Impaired ovalbumin-specific CD8 α ⁺ T-cell response by depletion of pDCs with 120G8 is shown in the bottom panel.

EMSA and immunoblot analysis

Electrophoretic mobility shift assay (EMSA) and immunoblot analyses were performed as described previously²². An oligonucleotide probe containing the NF- κ B-binding site of the IFN- β gene promoter was used. We performed supershift with the anti-RelA antibody (Santa Cruz). Immunoblotting was performed with anti-phospho-SAPK/JNK, anti-SAPK/JNK, anti-phospho-p38 MAPK, or anti-p38 MAPK antibodies (Cell Signalling Technology). The anti-IRF-7 polyclonal antibody was raised against the peptide STVGPATENREEVSLs by rabbit immunization.

Viral infections

Virus stocks were grown and virus titres quantified as described previously⁷. MEFs were infected with HSV-1 (1 multiplicity of infection (M.O.I.)), VSV (1 M.O.I.) or EMCV (0.1 M.O.I.). Dendritic cells were infected with HSV-1 (10 M.O.I.) or VSV (10 M.O.I.). For the *in vivo* studies, mice were intravenously infected with 1×10^7 plaque-forming units of HSV-1 or intraperitoneally infected with 1×10^5 plaque-forming units of EMCV, and sera were collected 12 h after infection. Levels of IFN- α were monitored by ELISA.

Antigen-specific CD8⁺ T-cell response

The visualization of ovalbumin-specific CD8⁺ T cells by major histocompatibility complex (MHC) class I tetramer was carried out as described previously²⁸. Briefly, mice were intraperitoneally injected with 0.5 mg of whole ovalbumin (Sigma-Aldrich) and 50 μ g of anti-CD40 (clone FGK45; Alexis) with or without 25 μ g of MALP2 (Alexis) or 50 μ g of CpG-A complexed to DOTAP. Six days later, spleens of the mice were removed and homogenized into single-cell suspensions. The cells were triply stained with fluorescein isothiocyanate-conjugated anti-CD8 α antibody, APC-conjugated CD44 antibody and phycoerythrin-conjugated K^b/ovalbumin tetramer according to the manufacturer's instructions (MBL). Cells were then analysed using a BD Biosciences FACS Calibur flow cytometer. For the pDC depletion experiment, mice were injected intraperitoneally with 1 mg of purified 120G8 antibody²⁹ or control rat IgG (Jackson ImmunoResearch) at day 1 and day 0 before immunization.

Received 4 January; accepted 14 February 2005; doi:10.1038/nature03464.
Published online 30 March 2005.

1. Taniguchi, T., Ogasawara, K., Takaoka, A. & Tanaka, N. IRF family of transcription factors as regulators of host defense. *Annu. Rev. Immunol.* **19**, 623–655 (2001).
2. Takeda, K. & Akira, S. Toll-like receptors in innate immunity. *Int. Immunol.* **17**, 1–14 (2005).
3. Wagner, H. The immunobiology of the TLR9 subfamily. *Trends Immunol.* **25**, 381–386 (2004).
4. Colonna, M., Trinchieri, G. & Liu, Y. J. Plasmacytoid dendritic cells in immunity. *Nature Immunol.* **5**, 1219–1226 (2004).
5. Sato, M. *et al.* Positive feedback regulation of type I IFN genes by the IFN-inducible transcription factor IRF-7. *FEBS Lett.* **441**, 106–110 (1998).
6. Marie, I., Durbin, J. E. & Levy, D. E. Differential viral induction of distinct interferon- α genes by positive feedback through interferon regulatory factor-7. *EMBO J.* **17**, 6660–6669 (1998).
7. Sato, M. *et al.* Distinct and essential roles of transcription factors IRF-3 and IRF-7 in response to viruses for IFN- α/β gene induction. *Immunity* **13**, 539–548 (2000).
8. Diebold, S. S. *et al.* Viral infection switches non-plasmacytoid dendritic cells into high interferon producers. *Nature* **424**, 324–328 (2003).
9. Yoneyama, M. *et al.* The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nature Immunol.* **5**, 730–737 (2004).
10. Fitzgerald, K. A. *et al.* IKK ϵ and TBK1 are essential components of the IRF3 signaling pathway. *Nature Immunol.* **4**, 491–496 (2003).
11. Sharma, S. *et al.* Triggering the interferon antiviral response through an IKK-related pathway. *Science* **300**, 1148–1151 (2003).
12. Balachandran, S., Thomas, E. & Barber, G. N. A FADD-dependent innate immune mechanism in mammalian cells. *Nature* **432**, 401–405 (2004).
13. Krug, A. *et al.* Herpes simplex virus type 1 activates murine natural interferon-producing cells through toll-like receptor 9. *Blood* **103**, 1433–1437 (2004).
14. Diebold, S. S., Kaisho, T., Hemmi, H., Akira, S. & Reis e Sousa, C. Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. *Science* **303**, 1529–1531 (2004).
15. Heil, F. *et al.* Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. *Science* **303**, 1526–1529 (2004).
16. Lund, J. M. *et al.* Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proc. Natl Acad. Sci. USA* **101**, 5598–5603 (2004).
17. Hemmi, H., Kaisho, T., Takeda, K. & Akira, S. The roles of Toll-like receptor 9, MyD88, and DNA-dependent protein kinase catalytic subunit in the effects of two distinct CpG DNAs on dendritic cell subsets. *J. Immunol.* **170**, 3059–3064 (2003).
18. Honda, K. *et al.* Role of a transcriptional-transcriptional processor complex involving MyD88 and IRF-7 in Toll-like receptor signaling. *Proc. Natl Acad. Sci. USA* **101**, 15416–15421 (2004).
19. Kawai, T. *et al.* Interferon- α induction through Toll-like receptors involves a direct interaction of IRF7 with MyD88 and TRAF6. *Nature Immunol.* **5**, 1061–1068 (2004).
20. Verthelyi, D., Ishii, K. J., Gursel, M., Takeshita, F. & Klinman, D. M. Human peripheral blood cells differentially recognize and respond to two distinct CpG motifs. *J. Immunol.* **166**, 2372–2377 (2001).
21. Qing, J. *et al.* Transforming growth factor β /Smad3 signaling regulates IRF-7 function and transcriptional activation of the beta interferon promoter. *Mol. Cell Biol.* **24**, 1411–1425 (2004).
22. Takaoka, A. *et al.* Integral role of IRF-5 in the gene induction programme activated by Toll-like receptors. *Nature* **434**, 243–249 (2005).
23. Barchet, W. *et al.* Virus-induced interferon α production by a dendritic cell subset in the absence of feedback signaling *in vivo*. *J. Exp. Med.* **195**, 507–516 (2002).
24. Krieg, A. M. *et al.* CpG motifs in bacterial DNA trigger direct B-cell activation. *Nature* **374**, 546–549 (1995).
25. Hochrein, H. *et al.* Herpes simplex virus type-1 induces IFN- α production via Toll-like receptor 9-dependent and -independent pathways. *Proc. Natl Acad. Sci. USA* **101**, 11416–11421 (2004).
26. Wagner, H., Heit, A., Schmitz, F. & Bauer, S. Targeting split vaccines to the endosome improves vaccination. *Curr. Opin. Biotechnol.* **15**, 538–542 (2004).

27. Klinman, D. M. Immunotherapeutic uses of CpG oligodeoxynucleotides. *Nature Rev. Immunol.* **4**, 249–258 (2004).
28. Ahonen, C. L. *et al.* Combined TLR and CD40 triggering induces potent CD8⁺ T cell expansion with variable dependence on type I IFN. *J. Exp. Med.* **199**, 775–784 (2004).
29. Asselin-Paturel, C., Brizard, G., Pin, J. J., Briere, F. & Trinchieri, G. Mouse strain differences in plasmacytoid dendritic cell frequency and function revealed by a novel monoclonal antibody. *J. Immunol.* **171**, 6466–6477 (2003).
30. Le Bon, A. *et al.* Cross-priming of CD8⁺ T cells stimulated by virus-induced type I interferon. *Nature Immunol.* **4**, 1009–1015 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Vilecek, H. Rosen, H. Ohno, F. Nakatsu, A. Nakano, L. Cantley, T. Saito, T. Seya, W.-C. Yeh and M. Lamphier for advice; S. Akira for MyD88 mutant mice; K. Yokote for Smad3 mutant mice; G. Trinchieri for pDC-specific antibody; and M. Shishido for technical assistance. This work was supported in part by a grant for Advanced Research on Cancer and a Grant-In-Aid for Scientific Research on Priority Areas from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, Uehara Memorial Foundation, the Sumitomo Foundation, and the Nakajima Foundation. H.Y. is a research fellow of the Japan Society for the Promotion of Science. H.N. was supported by an Ishidu Shun Memorial Scholarship.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.T. (tada@m.u-tokyo.ac.jp).

Self-organized patchiness in asthma as a prelude to catastrophic shifts

Jose G. Venegas^{1,2*}, Tilo Winkler^{1,2*}, Guido Musch^{1,2},
Marcos F. Vidal Melo^{1,2}, Dominick Layfield¹, Nora Tgavalekos^{1,3},
Alan J. Fischman^{1,2}, Ronald J. Callahan^{1,2}, Giacomo Bellani¹
& R. Scott Harris^{1,2}

¹Massachusetts General Hospital, Departments of Anesthesia and Critical Care, Radiology, and Medicine (Pulmonary and Critical Care Unit), 55 Fruit Street, Boston, Massachusetts 02114, USA

²Harvard Medical School, 25 Shattuck Street, Boston, Massachusetts 02115, USA

³Boston University, 1 Sherborn Street, Boston, Massachusetts 02215, USA

* These authors contributed equally to this work

Asthma is a common disease affecting an increasing number of children throughout the world. In asthma, pulmonary airways narrow in response to contraction of surrounding smooth muscle. The precise nature of functional changes during an acute asthma attack is unclear. The tree structure of the pulmonary airways has been linked to complex behaviour in sudden airway narrowing^{1,2} and avalanche-like reopening^{3,4}. Here we present experimental evidence that bronchoconstriction leads to patchiness in lung ventilation, as well as a computational model that provides interpretation of the experimental data. Using positron emission tomography, we observe that bronchoconstricted asthmatics develop regions of poorly ventilated lung. Using the computational model we show that, even for uniform smooth muscle activation of a symmetric bronchial tree, the presence of minimal heterogeneity breaks the symmetry and leads to large clusters of poorly ventilated lung units. These clusters are generated by interaction of short- and long-range feedback mechanisms, which lead to catastrophic shifts similar to those linked to self-organized patchiness in nature^{5,6}. This work might have implications for the treatment of asthma, and might provide a model for studying diseases of other distributed organs.

In the human lung, air is distributed into more than 130,000 terminal bronchioles by 16 generations of a dichotomous bronchial tree. Each terminal bronchiole further subdivides six more times

into thin walled alveolar ducts and alveoli, where gas exchange between capillary blood and alveolar gas takes place. During an asthma attack, inflammation, secretions, and the shortening of muscle fibres around the bronchial walls all cause narrowing of the airways, which obstructs airflow, causes heterogeneous ventilation, impairs gas exchange and makes breathing difficult. Although all branches of the tree up to the terminal bronchioles constrict, both experimental⁷ and theoretical⁸ evidence suggest that most of the breathing difficulty in asthma is caused by severe constriction of the terminal bronchioles. One might expect that constriction of these small airways, if independent and randomly distributed, should present a diffuse pattern of ventilation impairment throughout the lungs. However, magnetic resonance images of asthmatics^{9,10} show the presence of 'ventilation defects' much larger in size than expected for independent and randomly distributed constricted terminal bronchioles—a finding interpreted as the result of obstructing large airways. Although these ventilation defects could also be formed by clustered groups of constricted terminal bronchioles, magnetic resonance imaging cannot confirm or refute this hypothesis. To elucidate whether clusters of constricted small airways could be consistent with the observed ventilation defects, we used positron emission tomography (PET) to map the topographic distribution of ventilation in bronchoconstricted asthmatics (see Methods).

Consistent with previous magnetic resonance studies, our PET images show large ventilation defects in these bronchoconstricted asthmatics. Such ventilation defects are usually, but not always,

located in gravity-dependent regions of the lung (Fig. 1a, b and Supplementary Fig. S5). Quantitative analysis reveals systematic differences between lung units inside and outside the ventilation defects (Fig. 1c). Ventilation histograms for units inside the defects are bimodal and include a substantial fraction (>0.8) of the very low ventilating units of the lung (units receiving less than 0.5% of the mean lung unit ventilation). However, dispersed throughout the ventilation defects are units with normal ventilation levels. Lung units outside the defects have, on average, higher ventilation than units inside the defects, and contain only a small fraction of the lung units with very low ventilation. It is important to note that the bimodality of ventilation inside the ventilation defects is largely the result of heterogeneity at a small scale (at length scales <13 mm), as previously seen in bronchoconstricted sheep¹¹. PET imaging also shows that deep inhalations enhance non-uniform tracer removal from localized regions inside the defects. If severe constriction of large airways were the sole cause of these large ventilation defects, tracer washout and ventilation would have been unimodal (and low) throughout the defects, whereas reopening of a single constricted airway with deep inhalations would have caused uniform enhancement of tracer removal from the defects. It follows that the imaging data are not consistent with the exclusive constriction of large airways, but could instead be the result of clusters of constricted small airways.

To explore whether the imaging observations could be explained by self-organized clusters of constricted terminal bronchioles, we formulated a network model of ventilation distribution by incorporating into a tree structure a recent model of a single terminal airway^{12,13}. In the lungs, airways are embedded in parenchyma and thus their walls are tethered by alveolar septa, forming a network under tension. The single terminal airway model^{12,13} describes the local interplay between parenchymal tethering forces, intra- and extra-luminal pressures, and smooth muscle forces on a terminal bronchiole. During inhalation, tethering forces increase, and this tends to expand the imbedded airways. As smooth muscle constricts, reducing the airway lumen, the volume of air flowing into the bronchiole is also reduced. This lowers the tethering/expanding force, and makes the airway smaller, further decreasing the air flowing into it. This local interaction yields a multi-valued relationship between driving pressure and airway size that allows two stable states: one where the airway is nearly closed, and airflow, tidal volume and lung tissue tethering forces are greatly reduced, and another state where the airway is kept open by increased lung tissue tethering forces^{12,13}. In the network model described here, equivalent tissue, smooth muscle and gas pressure forces are estimated for each branch of a symmetric human bronchial tree model (see Methods). The model is numerically solved to yield a spatial distribution of ventilation into the tree's terminal units during simulated breathing. For illustration purposes, the spatial distribution of ventilation is displayed in a colour scale on a 64×64 grid of terminal units at the end of a Mandelbrot-like tree¹⁴ (Fig. 2a).

Model simulations involve uniform activation of smooth muscle throughout the airway tree, while tidal (breathing) volume and frequency are kept constant until a steady state is reached. To break the numerical symmetry of this uniform system, we introduce a small random heterogeneity (1% coefficient of variation) to airway wall thickness. Equivalent model predictions are obtained when the small heterogeneity is added to any other structural or functional parameter of the model. At low smooth muscle activation, the ventilation displayed by the model is rather uniform, as expected for a virtually uniform tree structure (see Supplementary Video S6). However, once smooth muscle activation reaches a critical level, a cluster of poorly ventilated terminal airways forms abruptly and expands in discrete steps, showing rich bifurcation dynamics as smooth muscle activation is further increased (Supplementary Fig. S7). The clustered pattern of constriction is not the result of a specific, small random heterogeneity imposed on airway wall

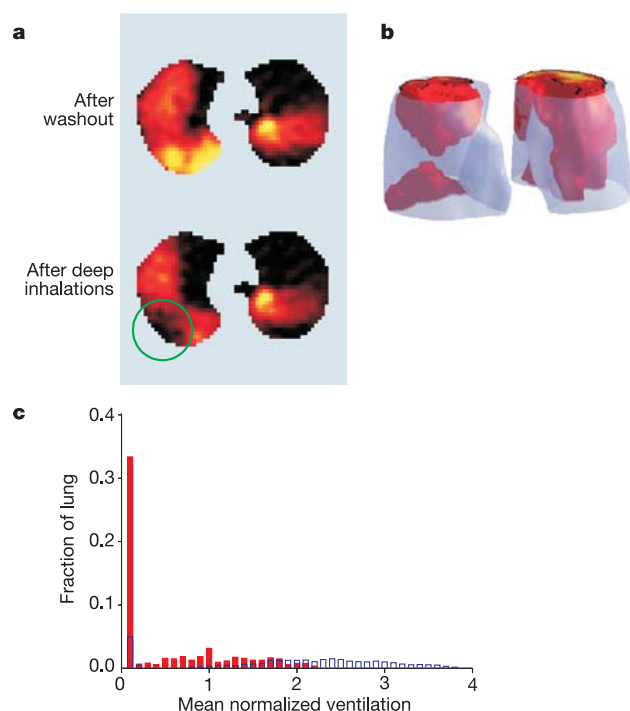


Figure 1 Heterogeneity in bronchoconstriction of an asthmatic's lung. **a**, Residual intrapulmonary ^{13}N tracer gas activity in a representative lung cross-section after intravenous bolus injection of ^{13}N -saline solution. Tracer concentration increases according to the following colour scale: black (no tracer), red, yellow, and white (highest). The insoluble tracer is washed out during breathing or was retained inside large ventilation defects. After deep inhalations (lower panel), tracer clearance is enhanced from parts of these defects (circle). **b**, Volumetric rendering of ventilation defects (red) and the external surface of the lungs (blue); image orientation is as if the subject were standing facing the reader. **c**, Histograms of mean normalized regional lung ventilation across differentiating units inside (red) and outside (blue) of the ventilation defects.

thickness. Similar results are obtained for different seeds that generate random heterogeneity. The system is therefore critical in the sense that small variations in its properties create very large effects.

Histograms of ventilation for units inside these clusters are bimodal (Fig. 2b), and include very poorly ventilated units interspersed with normally ventilating units (Fig. 2a), similar to those measured inside the ventilation defects of the bronchoconstricted lung. Ventilation outside these clusters was virtually uniform and included no units with very low ventilation. The area covered by clusters and the number of poorly ventilated units inside that area are inversely affected by tidal (breathing) volume and are history-dependent (Fig. 3 and Supplementary Fig. S8). It should be noted that in this network model, constriction of airways takes place at all levels of the tree but is clearly exaggerated in airways leading to a cluster, and it is highest in those terminal bronchioles leading to poorly ventilated terminal units (Supplementary Fig. S9).

The contribution of the branching tree to the ventilation distribution pattern is also revealed by conducting model simulations under three different conditions (Fig. 4): (1) with small heterogeneity and constriction allowed throughout the tree, (2) with constriction allowed throughout the tree but small heterogeneity limited to the terminal units, or (3) with constriction and heterogeneity limited to the terminal units. These simulations demonstrate that clusters form only for conditions (1) and (2). As in the PET images, the distribution of ventilation inside the clusters is bimodal, including all of the poorly ventilated units as well as a second mode of units ventilating at lower levels than those of units outside the cluster (histograms for conditions (1) and (2) are similar to that in Fig. 2b). If airway constriction is limited to terminal units (condition (3)) the ventilation in the network model is still bimodal, but the lack of constriction in the rest of the tree prevents spatial clustering (Fig. 4c). It is thus clear that constriction of the bronchial tree is necessary for the system to show self-organized clustering. Remarkably, for the same level of smooth muscle activation, the fraction of obstructed terminal units under condition (3) (0.08) is substantially less than those observed in conditions (1) (0.34) and (2) (0.35), where constriction of the tree above the terminal bronchioles is allowed.

The influence of airway tree structure on the spatial distribution of ventilation also becomes apparent by comparing the spatial pattern of ventilation inside and outside the clusters. The relationships between spatial heterogeneity of ventilation (assessed as a

coefficient of variation, $c.v. = \text{mean-normalized standard deviation}$) and the length scale (n) used for the measurement, obey power laws. In the absence of airway tree heterogeneity (Fig. 4b) or constriction (Fig. 4c), the exponents (m) of these relationships are consistent with the fractal dimension, D , expected for a purely random distribution ($D = 1 - m \approx 1.5$, ref. 15). In contrast, the effect of adding a small heterogeneity to the airway tree (Fig. 4a) is magnified, and ventilation distribution inside the cluster is consistent with a spatially correlated pattern with $D \approx 1.2$. This value of D falls within the observed range for pulmonary ventilation and perfusion in animals^{15,16}. Thus, the network model provides a functional link between microscopic- and organ-level behaviours that allows comparisons with imaging data beyond those possible with the previously reported single terminal bronchiole model.

Local bistable behaviour and short-range to long-range competing interactions result in the patchy self-organization postulated to precede catastrophic shifts in ecosystems⁶. Our network model and imaging data suggest that analogous mechanisms are involved in bronchoconstricted asthmatic lungs. It is well established that dynamic airway stretch increases mean airway calibre during breathing^{17,18} and that the absence of large periodic stretches promotes constriction¹⁹. In our model, we observe that short-range synergistic feedback is caused by propagation of constriction up and down the tree as a result of airway-tissue interactions that magnify small regional heterogeneities in the tree and contribute to break the symmetry of the system. For example, constriction of a large airway reduces flow to all the terminal units fed by that airway, causing a reduction in tissue expansion, which promotes constriction in all branches of the tree beyond that point and down to the terminal bronchioles. Similarly, slightly higher average constriction of a set of terminal bronchioles fed by a common tree branch reduces average tissue expansion of their common airway, which facilitates its constriction and, consequently, the constriction of other terminal units fed by that common branch. Note that unlike in ecosystems, this feedback does not depend on an absolute distance between lung units, but instead relies on a functional proximity through the airway tree (adjacent units on opposite sides of the centre line of the model are far apart because they only share the largest airway of the model).

Long-range competitive feedback in the lung is caused by airflow redistribution away from constricting regions, which increases expanding tissue forces and promotes bronchial dilatation in the rest of the lung, thus protecting it from catastrophic closure. Again

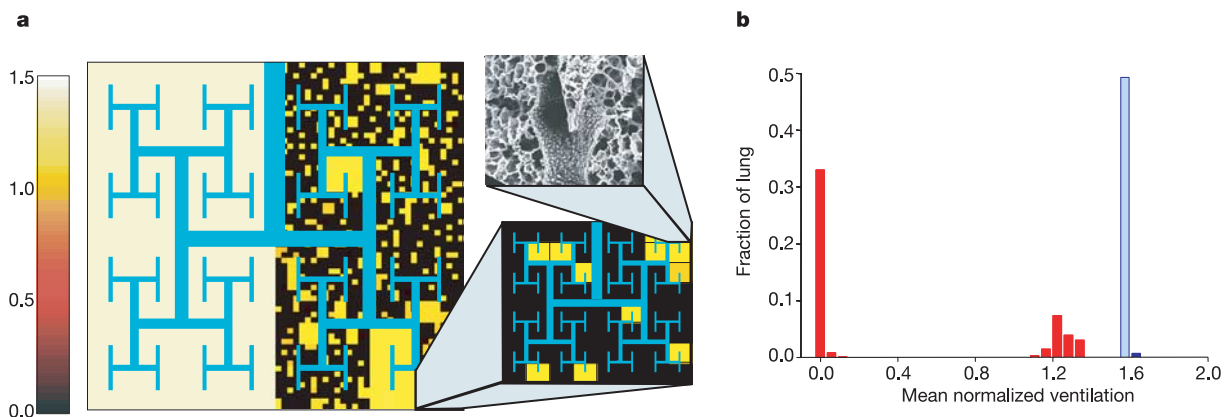


Figure 2 Modelling of bronchoconstricted lungs. **a**, Schematic two-dimensional representation of a Mandelbrot-like bronchial tree¹⁴ where the background shows (in colour scale) the distribution of mean normalized ventilation into its terminal branches. Top inset, a photograph of a pulmonary bronchiole and tethering alveolar walls illustrates a terminal unit in the model. Bottom inset, zoom of the last seven generations of the

model's lower-right corner. The whole right side of the model tree shows a large cluster of severely obstructed terminal bronchioles (black). **b**, Histogram of mean normalized regional lung inside (red) and outside (blue) of the cluster (red). Note the bimodal histogram inside the cluster, with a large fraction of poorly ventilated lung units.

unlike ecosystems, this feedback is not a function of physical distance, because it affects all the lung units that remain open. Nonetheless, the effect of the interaction between this distance-independent competitive feedback and the short-range synergistic feedback, as in ecosystems, is to promote the formation of clusters, which show catastrophic (stepwise) increases in size as tidal volume is reduced (Fig. 3) and/or as smooth muscle activation is increased (Supplementary Video S6). These mechanisms are also responsible for the regional bistability shown by the model in response to changes in tidal volume, such as the hysteresis in cluster size and location, and in the number of highly constricted terminal units. If we make the analogy between tissue distension in the lung and 'resource input' in ecosystems, the regional bistability shown in the lung model would be analogous to the global bistability proposed for ecosystems undergoing catastrophic shifts⁶.

How relevant are these results to the real lung, where the bronchial tree is asymmetric in structure and may have heterogeneous smooth muscle reactivity? Symmetry-breaking in the lung has been proposed as a source of heterogeneity in bronchoconstriction²⁰. Clearly, additional sources of heterogeneity, such as asymmetry by gravity-oriented gradients in lung expansion, should contribute to the preferential clustering of constricted airways in less expanded, gravity-dependent regions of the lung. Structural or functional heterogeneity in the tree would also explain why the

distributions of ventilation inside and outside the clusters seen in the PET imaging studies are broader than those shown in the homogeneous network model. Although the network model is based on the mechanical interdependence in expansion between airways and their surrounding parenchyma¹³, it neglects the interdependence in expansion between neighbouring airways and between neighbouring regions, and it also neglects heterogeneous lung expansion caused by active effort of chest wall muscles. Airway expansion interdependence would tend to expand airways near a constricted airway, but regional interdependence would tend to reduce expansion of regions near an expanded region. Differences in regional lung expansion, due to the motions of the diaphragm and the ribcage during inhalation, could also help to break symmetry. More advanced models are required to assess the quantitative relevance of these effects.

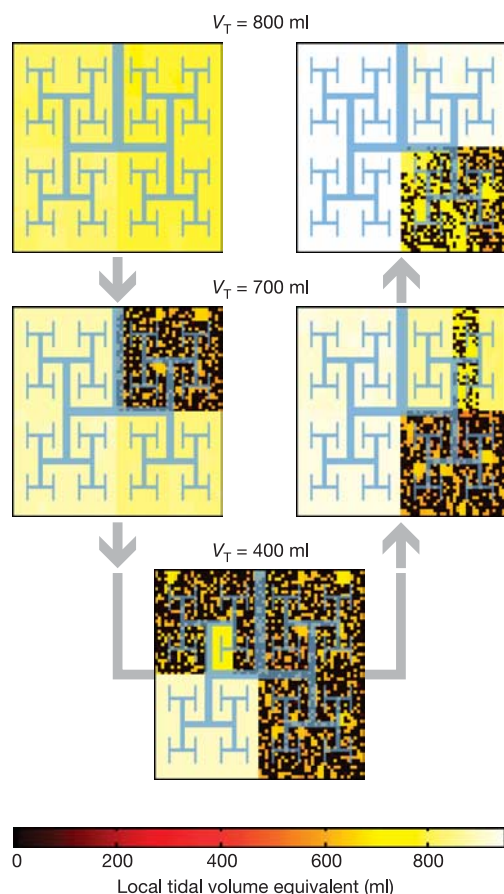


Figure 3 Tidal volume-dependent hysteresis in regional ventilation distribution to terminal units. This simulation was carried out at a constant degree of smooth muscle activation as tidal volume was decreased in steps from 800 ml to 700 ml then to 400 ml, and then increased back to 800 ml. Here the colour scale corresponds to the regional equivalent of a uniformly distributed tidal volume. Note the regional bistability and hysteresis shown by the model.

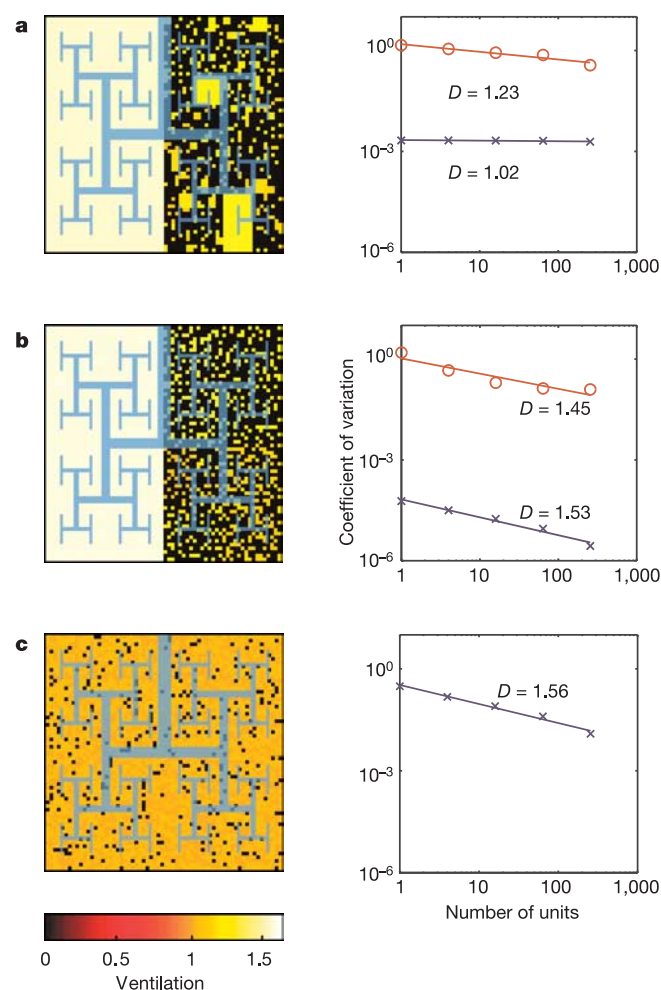


Figure 4 Distributions of ventilation under three conditions. Left panels show the conditions. **a**, Condition (1), with a small heterogeneity added and smooth muscle constriction allowed throughout the tree. **b**, Condition (2), with constriction allowed throughout the tree but small heterogeneity limited to terminal units. **c**, Condition (3), with constriction limited to the terminal units. Clusters form only for conditions (1) and (2). Remarkably, the fraction of severely obstructed terminal units in **c** (0.08) is substantially less than that observed in **a** (0.34) and **b** (0.35). Right panels show logarithmic plots (for each of the conditions above) of the coefficient of variation (c.v. = mean-normalized s.d. of ventilation) versus number of units averaged during the measurement. The relationships obey power laws consistent with a spatially correlated structure (fractal dimension $D \approx 1.2$) inside the cluster in **a**, with a uniform distribution ($D \approx 1.0$) outside the cluster in **a**, and with random distributions ($D \approx 1.5$) in **b** and **c**.

Qualitative magnetic resonance images after a single breath of hyperpolarized helium in mild asthmatics^{9,10,21} show large ventilation defects (corresponding to hundreds or thousands of terminal units), which reduce in number after administration of a bronchodilator⁹. High resolution computerized tomographic images taken after a maximal exhalation also show contiguous regions (of similar size) with trapped gas^{22,23}. The PET images presented here reveal both the presence of a significant number of units with good ventilation inside the ventilation defects, and a reduction in defect size after deep inhalations. The network model also shows co-existence of ventilated and poorly ventilated terminal units inside a cluster, as well as cluster size reduction with larger tidal volumes. Furthermore, the model predicts that the degree of peripheral constriction can be substantially reduced by reducing constriction of central airways (Fig. 4c), a finding that could provide an explanation for the effectiveness of novel asthma treatments that target the contractility of large airways^{24,25}.

The simulations of this model were conducted while forcing a constant tidal volume into the lung, independent of the constriction level. This is a reasonable assumption for mild degrees of bronchoconstriction, when the asthmatic is capable of maintaining a normal ventilation level. The simulations also suggest that in severe asthma, when increased respiratory effort and fatigue reduce tidal volume and parenchymal expansion, the remaining open part of the lung will be exposed to catastrophic airway closures—a process that without treatment can rapidly lead to respiratory failure.

Clustered bronchoconstriction could also limit the effectiveness of inhaled therapeutic drugs for asthma by promoting delivery to regions least in need of medication. Indeed, delivering bronchodilation medication to well-ventilated regions could redirect even more ventilation to them, further reducing lung expansion of constricted areas and exacerbating ventilation heterogeneity. This could explain the ineffectiveness of inhaled bronchodilators in some patients, as well as the need to deliver the drug through the vascular system in some cases of severe asthma²⁶. We conclude that the potential for catastrophic shifts in airway constriction, and the clustered distribution of obstruction, might be involved in sudden and unexplained asthma attacks that are consistently difficult to treat. In the same way that patchiness might indicate a propensity towards catastrophic shifts in ecosystems, self-organized clustering in bronchoconstriction might be the prelude to dangerous respiratory failure in asthma. □

Methods

PET imaging of regional ventilation

We studied mild and moderate asthmatics during bronchoconstriction provoked by inhalation of sufficient methacholine aerosol to reduce by 20% the volume exhaled during the first second of forced exhalation from a maximally inflated lung. A series of dynamic PET images, consisting of 15 contiguous, 6.5-mm-thick cross-sectional slices covering 10 cm of the chest, was acquired following an intravenous bolus injection of the positron-emitter nitrogen-13 (¹³NN), where ¹³NN is a molecule of N₂ gas in which one atom is ¹³N and the other is ¹⁴N in saline solution²⁷. The solubility of ¹³NN in water and blood is low (partition coefficient ~0.014). Thus, upon arrival at the pulmonary capillary bed during a short breath-hold, ¹³NN diffuses into the alveolar gas before it washes out of the lungs during breathing (Supplementary video S3 and S4). At the end of a two-minute washout period, the subject was asked to take three consecutive deep breaths; PET imaging was then continued for an additional minute. Quantitative analysis of these images involves modelling the local pulmonary ¹³NN tracer kinetics to derive topographic ventilation and perfusion data (including sub-resolution heterogeneity), and to form histograms of ventilation distribution¹¹.

Network model of bronchoconstriction

We began with twelve generations of a symmetric dichotomous and fully relaxed airway tree with dimensions equal to those from the fourth to the sixteenth generations of the human airway tree²⁸. The smallest airways of the model correspond to terminal bronchioles that feed 4,096 independent elastic elements residing within a single elastic thorax. For each branch of the tree, airflow and pressure losses are calculated assuming laminar flow and neglecting the effects of gas compression and dynamic airway calibre fluctuations during breathing. Airway wall thickness is adjusted to preserve wall volume during constriction while airway length is kept constant²⁹. Each airway is surrounded by the lung tissue that it feeds. The tissue and pressure forces acting on an airway wall are

calculated as a function both of tissue mechanical properties and of average expansion of the terminal units subtended by that airway¹³. Mechanical properties of the airway wall are derived from the behaviour of activated smooth muscle fibres submitted to periodic stretching³⁰. The properties at variable levels of smooth muscle activation are linearly scaled from the muscle's tension-length relationship at maximal activation¹³. Airway radius is estimated from those properties and the peak transmural pressure to which the airway wall is exposed during a breath¹³.

For the 8,191 tree branches, recursive equations are formulated to relate internal diameter, length, instantaneous luminal pressure and flow to tissue/airway forces during breathing (see Supplementary Methods). To simplify the computations, equations are solved for conditions of cyclic breathing with steady inspiratory flow rate and passive exhalation. Tidal (breath) volume into the first branch of the tree is constrained to be constant. Airway diameters are iteratively adjusted to converge into steady-state values consistent with both the mechanics of the airway walls and the airflow distribution within the tree.

Spatial heterogeneity of ventilation was assessed as a coefficient of variation (c.v. = mean-normalized standard deviation). To test whether the spatial distributions within and outside the clusters were consistent with fractal distributions, we plotted (Fig. 4) the c.v. against the length scale used for the measurement (N = number of units averaged in the measurement of c.v.). The exponents (m) of these power law relationships were used to estimate the fractal dimension ($D = 1 - m$; refs 15, 16).

Received 16 December 2004; accepted 7 February 2005; doi:10.1038/nature03490.

Published online 16 March 2005.

- Suki, B. & Frey, U. Temporal dynamics of recurrent airway symptoms and cellular random walk. *J. Appl. Physiol.* **95**, 2122–2127 (2003).
- Mauroy, B., Filoche, M., Weibel, E. R. & Sapoval, B. An optimal bronchial tree may be dangerous. *Nature* **427**, 633–636 (2004).
- Suki, B., Barabasi, A. L., Hantos, Z., Petak, F. & Stanley, H. E. Avalanches and power-law behaviour in lung inflation. *Nature* **368**, 615–618 (1994).
- Alencar, A. M. et al. Physiology: Dynamic instabilities in the inflating lung. *Nature* **417**, 809–811 (2002).
- Judd, S. L. & Silber, M. Simple and superlattice Turing patterns in reaction-diffusion systems: bifurcation, bistability, and parameter collapse. *Physica D* **136**, 45–65 (2000).
- Rietkerk, M., Dekker, S. C., de Ruiter, P. C. & van de Koppel, J. Self-organized patchiness and catastrophic shifts in ecosystems. *Science* **305**, 1926–1929 (2004).
- Beigelman-Aubry, C. et al. Mild intermittent asthma: CT assessment of bronchial cross-sectional area and lung attenuation at controlled lung volume. *Radiology* **223**, 181–187 (2002).
- Lutchen, K. R. et al. Airway constriction pattern is a central component of asthma severity: the role of deep inspirations. *Am. J. Respir. Crit. Care Med.* **164**, 207–215 (2001).
- Samee, S. et al. Imaging the lungs in asthmatic patients by using hyperpolarized helium-3 magnetic resonance: assessment of response to methacholine and exercise challenge. *J. Allergy Clin. Immunol.* **111**, 1205–1211 (2003).
- Klarreich, E. Take a deep breath. *Nature* **424**, 873–874 (2003).
- Vidal Melo, M. F. et al. Quantification of regional ventilation-perfusion ratios with PET. *J. Nucl. Med.* **44**, 1982–1991 (2003).
- Anafi, R. C., Beck, K. C. & Wilson, T. A. Impedance, gas mixing, and bimodal ventilation in constricted lungs. *J. Appl. Physiol.* **94**, 1003–1011 (2003).
- Anafi, R. C. & Wilson, T. A. Airway stability and heterogeneity in the constricted lung. *J. Appl. Physiol.* **91**, 1185–1192 (2001).
- Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures* 71–73 (Princeton Univ. Press, Princeton, 1991).
- Venegas, J. G. & Galletti, G. G. Low-pass filtering, a new method of fractal analysis: application to PET images of pulmonary blood flow. *J. Appl. Physiol.* **88**, 1365–1373 (2000).
- Altemeier, W. A., McKinney, S. & Glenny, R. W. Fractal nature of regional ventilation distribution. *J. Appl. Physiol.* **88**, 1551–1557 (2000).
- Fredberg, J. J. Bronchospasm and its biophysical basis in airway smooth muscle. *Respir. Res.* **5**, 2 (2004).
- Latourelle, J., Fabry, B. & Fredberg, J. J. Dynamic equilibration of airway smooth muscle contraction during physiological loading. *J. Appl. Physiol.* **92**, 771–779 (2002).
- Skloot, G., Permutt, S. & Togias, A. Airway hyperresponsiveness in asthma: a problem of limited smooth muscle relaxation with inspiration. *J. Clin. Invest.* **96**, 2393–2403 (1995).
- Bates, J. H. Bronchoconstriction and broken symmetry. *Comments Theor. Biol.* **3**, 397–415 (1995).
- Wagers, S. Polarized helium: changing our view of asthma. *J. Allergy Clin. Immunol.* **111**, 1201–1202 (2003).
- Laurent, F., Latrabe, V., Raherison, C., Marthan, R. & Tunon-de-Lara, J. M. Functional significance of air trapping detected in moderate asthma. *Eur. Radiol.* **10**, 1404–1410 (2000).
- Newman, K. B., Lynch, D. A., Newman, L. S., Ellegood, D. & Newell, J. D. Jr Quantitative computed tomography detects air trapping due to asthma. *Chest* **106**, 105–109 (1994).
- Cox, P. G., Miller, J., Mitzner, W. & Leff, A. R. Radiofrequency ablation of airway smooth muscle for sustained treatment of asthma: preliminary investigations. *Eur. Respir. J.* **24**, 659–663 (2004).
- Danek, C. J. et al. Reduction in airway hyperresponsiveness to methacholine by the application of RF energy in dogs. *J. Appl. Physiol.* **97**, 1946–1953 (2004).
- Smith, D. et al. Intravenous epinephrine in life-threatening asthma. *Ann. Emerg. Med.* **41**, 706–711 (2003).
- Musch, G. et al. Topographical distribution of pulmonary perfusion and ventilation, assessed by PET in supine and prone humans. *J. Appl. Physiol.* **93**, 1841–1851 (2002).
- Weibel, E. R. *Morphometry of the Human Lung* 115–126 (Berlin, Springer, 1963).
- Smith, J. C., Butler, J. P. & Hopkin, F. G. Jr Contribution of tree structures in the lung to lung elastic recoil. *J. Appl. Physiol.* **57**, 1422–1429 (1984).
- Gunst, S. J. & Stropp, J. Q. Pressure-volume and length-stress relationships in canine bronchi *in vitro*. *J. Appl. Physiol.* **64**, 2522–2531 (1988).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to T. A. Wilson, K. M. Miller, E. Lowenstein, B. Suki, J. P. Butler, M. Reiterkerk and J. J. Fredberg for their suggestions and comments during the preparation of the manuscript. J. A. Correia and W. M. Bucelewicz are thanked for their help with technical aspects of the tracer preparation. This work was funded by an NIH HLBI grant.

Authors' contributions J.G.V. and T.W. contributed equally to the theoretical aspects of this work. All authors contributed to experimental components of this work.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.G.V. (jvenegas@vqpet.mgh.harvard.edu).

Low dose oral cannabinoid therapy reduces progression of atherosclerosis in mice

Sabine Steffens¹, Niels R. Veillard^{1*}, Claire Arnaud^{1*}, Graziano Pelli¹, Fabienne Burger¹, Christian Staub³, Andreas Zimmer⁴, Jean-Louis Frossard² & François Mach¹

¹Division of Cardiology, Department of Medicine, Foundation for Medical Research,

²Division of Gastroenterology and ³Institute of Legal Medicine, University Hospital, Faculty of Medicine, 1211 Geneva, Switzerland

⁴Laboratory for Molecular Neurobiology, Department of Psychiatry, University of Bonn, 53105 Bonn, Germany

* These authors contributed equally to this work

Atherosclerosis is a chronic inflammatory disease, and is the primary cause of heart disease and stroke in Western countries¹. Derivatives of cannabinoids such as delta-9-tetrahydrocannabinol (THC) modulate immune functions² and therefore have potential for the treatment of inflammatory diseases. We investigated the effects of THC in a murine model of established atherosclerosis. Oral administration of THC (1 mg kg⁻¹ per day) resulted in significant inhibition of disease progression. This effective dose is lower than the dose usually associated with psychotropic effects of THC. Furthermore, we detected the CB2 receptor (the main cannabinoid receptor expressed on immune cells^{2,3}) in both human and mouse atherosclerotic plaques. Lymphoid cells isolated from THC-treated mice showed diminished proliferation capacity and decreased interferon- γ secretion. Macrophage chemotaxis, which is a crucial step for the development of atherosclerosis¹, was also inhibited *in vitro* by THC. All these effects were completely blocked by a specific CB2 receptor antagonist⁴. Our data demonstrate that oral treatment with a low dose of THC inhibits atherosclerosis progression in the apolipoprotein E knockout mouse model, through pleiotropic immunomodulatory effects on lymphoid and myeloid cells. Thus, THC or cannabinoids with activity at the CB2 receptor may be valuable targets for treating atherosclerosis.

It is now generally recognized that atherosclerosis is a chronic inflammatory disease that can lead to acute clinical events following plaque rupture and thrombosis^{1,5}. Current treatments for atherosclerosis are mainly based on drugs that lower plasma cholesterol concentration and blood pressure. In particular, statins have proven to reduce cardiovascular events significantly, not only as a consequence of their cholesterol-lowering properties but also through their more recently identified anti-inflammatory and immunomodulatory effects⁶. Nevertheless, atherosclerosis remains the primary cause of heart disease and stroke in Western countries, accounting for up to 50% of deaths. The identification and development of

promising new anti-inflammatory therapies is therefore of great medical interest. Immunosuppressive and anti-inflammatory effects of cannabinoids have been reported^{7–9}, and pre-clinical studies have provided therapeutic rationale for their use in treating autoimmune diseases such as multiple sclerosis¹⁰ and rheumatoid arthritis¹¹. In murine collagen-induced arthritis (a mouse model of rheumatoid arthritis), cannabidiol, a major cannabinoid derivative, ameliorates chronic inflammation by inhibiting T-helper type 1 (T_H1) responses, as shown by reduced proliferation and interferon- γ (IFN- γ) production by lymph node cells from treated mice¹¹. In support of the immunomodulatory role of cannabinoids, receptors for THC have been identified on several types of immune cells. The CB1 receptor is expressed predominantly in the brain, but CB2 receptor expression is found primarily on cells of the immune system, including B cells, T cells and monocytes². It has been suggested that the immunomodulatory effects of cannabinoids are mediated by the CB2 receptor expressed on immune cells³. The fact that THC-mediated inhibition of T-helper cell activation is not

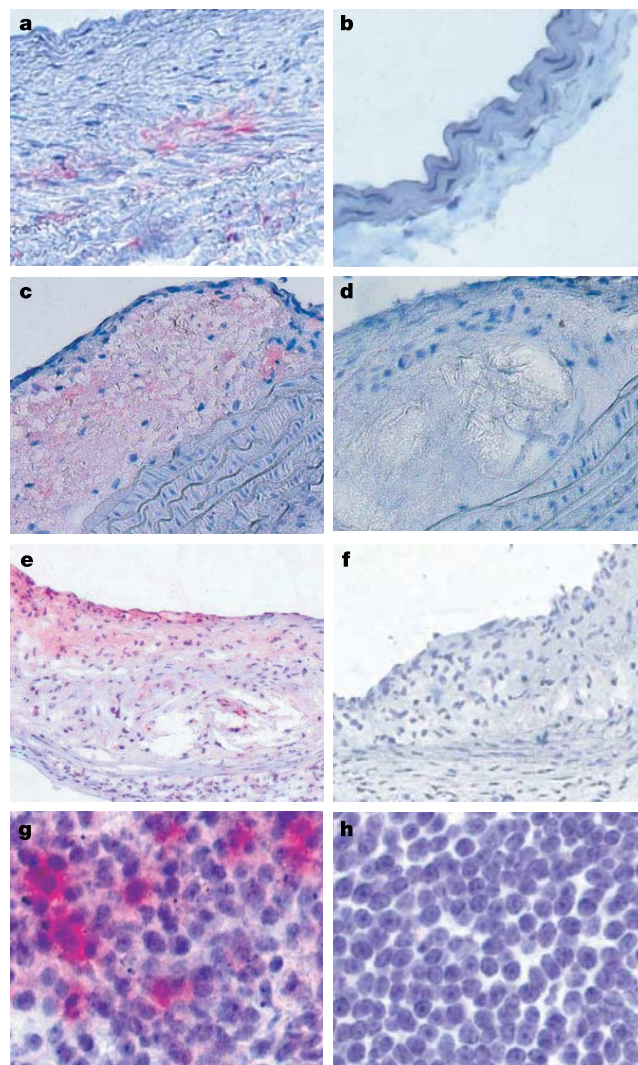


Figure 1 The cannabinoid receptor CB2 is expressed in human and mouse atherosclerotic plaques. **a–h**, Representative cryosections showing CB2 receptor expression (pink staining) in human coronary atherosclerotic lesion (**a**), normal carotid artery from wild-type mouse (**b**), aortic arch atherosclerotic lesion from ApoE^{-/-} mouse (**c**, **d**), aortic root atherosclerotic lesion from ApoE^{-/-} mouse (**e**, **f**), and spleen (follicular area) from wild-type (**g**) or CB2^{-/-} mouse (**h**). Sections were immunolabelled with an anti-CB2 receptor antibody (**a–c**, **e**, **g**, **h**), or with secondary antibody only (**d**, **f**).

observed in CB2 receptor knockout mice strongly supports this hypothesis³. These immunomodulatory properties suggest that cannabinoid derivatives might be beneficial in the treatment of atherosclerosis.

We hypothesized that cannabinoid treatment would alter inflammatory processes pivotal for the development of atherosclerosis, thus limiting disease progression. We first analysed the expression of CB2 receptors in atherosclerotic plaques of human and mouse diseased arteries. Immunohistochemistry revealed the presence of the CB2 receptor within human coronary atheroma as well as in atherosclerotic lesions of mouse aortic arch and root, but no CB2 staining was observed in non-diseased arteries (Fig. 1). CB1 receptors were not detected in any vascular tissue (data not shown). Double immunofluorescence staining confirmed that CB2 receptors are expressed by macrophages and T lymphocytes within atherosclerotic lesions (Fig. 2).

The anti-atherosclerotic potential of THC was tested in the apolipoprotein E knockout (ApoE^{-/-}) mouse model of atherosclerosis. We chose THC, the major constituent of marijuana, because its immune modulating effects have been well documented^{7,9,10}. THC is already commercially available, for example as an anti-vomiting drug or for use in treating anorexia. Given that any new anti-inflammatory therapy should be well tolerated and preferably devoid of psychotropic effects, we first performed dose-response experiments of THC on early atherosclerotic lesion development. The maximal effect of THC on reduction of atherosclerotic lesions was observed at a dose of 1 mg kg⁻¹ per day, with lower effects at higher doses (data not shown). Similar effects of cannabinoid treatment have been observed in previous *in vivo* studies^{11,12}. However, the underlying mechanisms for these observations remain unclear. Analysis of THC levels in blood serum of THC-treated mice (1 mg kg⁻¹) showed a concentration of 0.6 ng ml⁻¹, which is lower than the dose usually associated with psychotropic effects in either humans or mice¹³⁻¹⁶. THC is known to be very lipid-soluble and is stored in fat tissue, which might explain the low THC levels observed in the serum of hypercholesterolemic mice. As the ApoE^{-/-} mouse model used in our experiments shows

significant accumulation of fat tissue, especially within the vessel wall, we might speculate that active THC is stored in the atherosclerotic plaque. In addition, circulating THC could be bound to the vast numbers of circulating lipid particles. Given the unique lipid metabolism of ApoE^{-/-} mice and the high liposolubility of THC, ApoE^{-/-} mice might be uniquely sensitive to very low doses of THC.

On the basis of these results, we chose a dose of 1 mg kg⁻¹ THC for daily oral administration in order to test its therapeutic efficiency on advanced, established atherosclerosis. ApoE^{-/-} mice were fed with a high-cholesterol diet for 5 weeks, and THC was then administered daily for the next 6 weeks, while maintaining the cholesterol diet. After the initial 5 weeks on a high-cholesterol diet, atherosclerotic lesions were clearly detectable within the aortic roots of ApoE^{-/-} mice (Fig. 3a). More advanced vascular lesion development occurred in the aortic roots compared with the abdominal aorta, as demonstrated in previous studies¹⁷⁻¹⁹. After 11 weeks on a high-cholesterol diet, there was marked progression of atherosclerotic lesions within the aortic roots of control mice, but THC-treated mice showed significantly reduced progression of atherosclerotic lesions (Fig. 3a). Similar results were observed in the abdominal aorta (data not shown).

To determine whether this anti-atherosclerotic effect of THC was mediated through the CB2 receptor, we performed experiments using THC in the presence of the CB2 receptor-specific antagonist SR144528 (ref. 4). The inhibitory effect of THC on early atherosclerotic lesion progression was completely abolished in the presence of the CB2 receptor antagonist (Fig. 3b). There were no differences in serum cholesterol, triglyceride levels or body weights between the treatment groups. None of the THC-treated mice died during treatment or showed unhealthy behaviour. Quantitative immunostaining analysis of atherosclerotic lesions showed significantly reduced macrophage infiltration per lesion in THC-treated mice (Fig. 3c). The initial stages of cell recruitment into atherosclerotic lesions involve rolling and adhesion of leukocytes to the vessel wall. Intravital microscopy was used to visualize the effect of THC on leukocyte recruitment within the mesenteric microvascu-

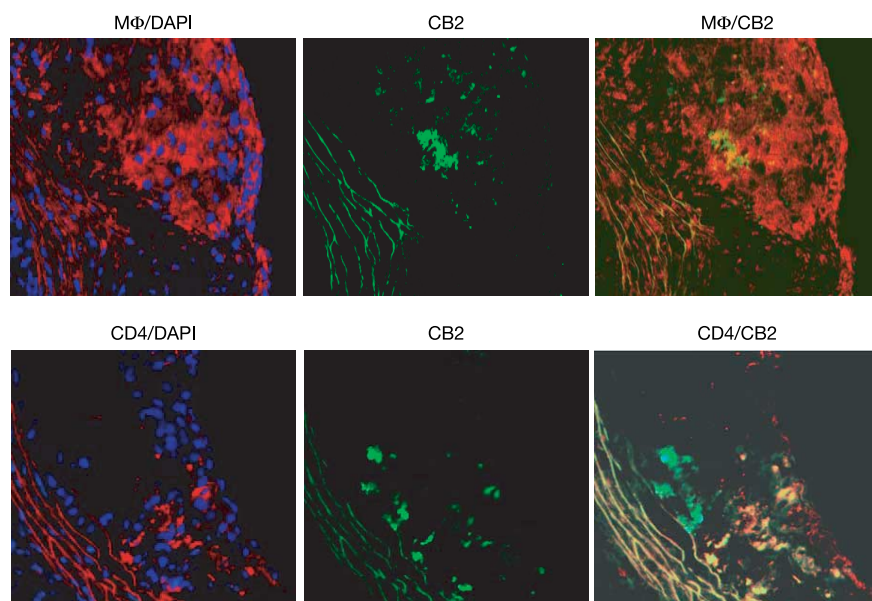


Figure 2 The cannabinoid receptor CB2 is expressed on macrophages and T lymphocytes within atherosclerotic plaques of ApoE^{-/-} mice. Shown are representative photomicrographs of ApoE^{-/-} mouse aortic root cryosections, with the lumen of the artery at the right side of each photo. Immunofluorescence double-labelling was

performed for CB2 receptors (green) and a cell-specific marker (red) for macrophages (MΦ) or T lymphocytes (CD4). Nuclei were stained in blue with DAPI. Analysis of atheroma from 4 different mice showed similar results.

lature²⁰. These experiments revealed significantly reduced leukocyte adhesion in THC-treated animals (1.82 ± 0.65 cells per $100 \mu\text{m}$) compared with controls (4.03 ± 0.44 cells per $100 \mu\text{m}$). The effect of THC on leukocyte adhesion was reversed when animals were also treated with the CB2 receptor antagonist SR144528 (4.28 ± 0.98 cells per $100 \mu\text{m}$; $P < 0.05$ for control versus THC and for THC versus THC + SR144528).

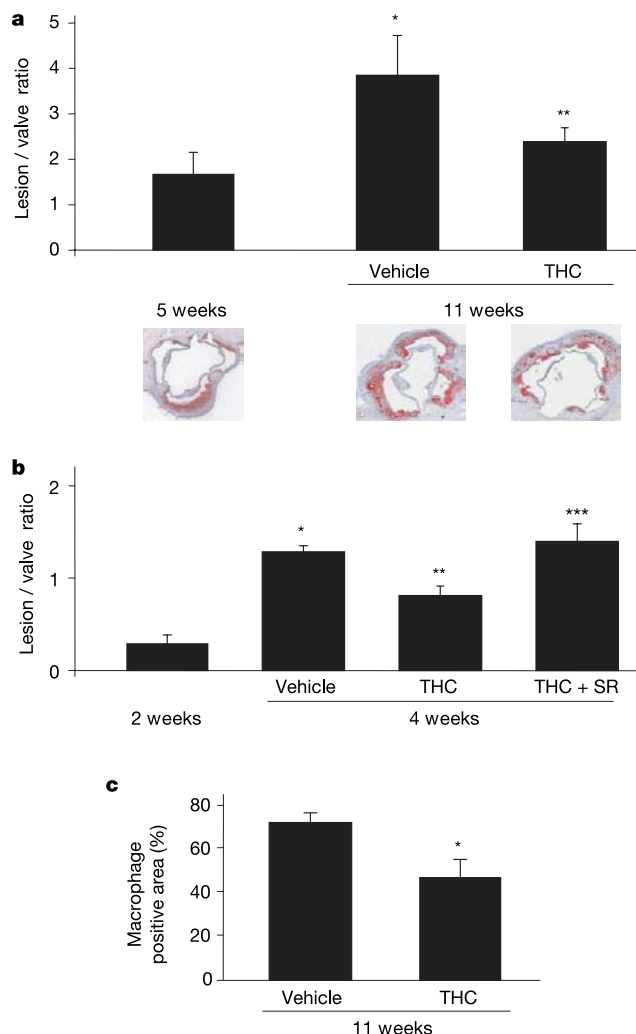


Figure 3 Reduced atherosclerotic plaque development and macrophage content in THC-treated ApoE^{-/-} mice. **a**, Representative cryosections of mouse aortic roots, stained with Sudan IV for lipid deposition, and quantification of atherosclerotic lesions. After 5 weeks on a high-cholesterol diet, ApoE^{-/-} mice developed atherosclerotic lesions ($n = 5$). THC (1 mg kg^{-1}) was orally administered during the last 6 weeks of the 11-week diet ($n = 6$ for THC-treated and $n = 8$ for control mice). Data represent mean values $\pm$ s.e.m. Single asterisk, $P < 0.05$ (compared with ApoE^{-/-} at 5 weeks); double asterisk, $P < 0.05$ (compared with ApoE^{-/-} at 11 weeks without THC). **b**, The CB2 receptor antagonist SR144528 abolishes the anti-atherosclerotic effect of THC. Atherosclerotic lesion development was detectable after 2 weeks on a high-cholesterol diet ($n = 4$). THC (1 mg kg^{-1}) alone or together with SR144528 was orally administered during the last 2 weeks of the 4-week diet, with control mice receiving corresponding dilutions of the SR144528 vehicle ($n = 6$ per group). Data represent mean values $\pm$ s.e.m. Single asterisk, $P < 0.05$ (compared with ApoE^{-/-} at 2 weeks); double asterisk, $P < 0.05$ (compared with ApoE^{-/-} at 4 weeks without THC); triple asterisk, $P < 0.05$ (compared with THC-treated ApoE^{-/-} mice). **c**, Reduced macrophage infiltration in atherosclerotic plaques from THC-treated mice after 11 weeks on a high-cholesterol diet. Quantification of labelling with an anti-macrophage antibody (macrophage-positive cells divided by the area of atherosclerotic lesion, $n = 6$ for both groups). Values represent mean values $\pm$ s.e.m.; single asterisk, $P < 0.05$.

The progression of atherosclerosis may result from an imbalance between pro- and anti-inflammatory mediators in response to endothelial injury²¹. Several reports demonstrate that T cells play an important role during early atherosclerosis development^{22,23}. It has been shown that T_H1 cells represent the predominant population of activated T cells within atherosclerotic lesions^{24,25}, and that THC treatment affects the T_H1/T_H2 balance of activated T cells^{8,9}. We hypothesized that the observed anti-atherosclerotic effects of THC might result from a modified cytokine expression pattern in the THC-treated mice. We therefore investigated the influence of THC treatment on inflammatory responses at the beginning of atherosclerosis development. We analysed proliferative responses and cytokine profiles of lymphoid cells isolated from mice on a high-cholesterol diet with and without administration of different doses of THC. Compared with untreated mice, treatment with 1 mg kg^{-1} THC significantly reduced proliferative responses of *in vitro* stimulated splenocytes, whereas both low (0.1 mg kg^{-1}) and high (10 mg kg^{-1}) doses had no significant effect (Fig. 4a). Similar results were obtained with lymph node cells (data not shown). We examined culture supernatants for T_H1 cytokines (IFN- γ , interleukin (IL)-12), T_H2 cytokines (IL-4, IL-10) and transforming growth factor (TGF)- β . Similar to the proliferative response, only the 1 mg kg^{-1} THC dose significantly reduced IFN- γ compared with controls. However, this dose resulted in only a modest, non-significant downregulation of IL-10 and TGF- β (Fig. 4b–d). In both THC-treated and untreated groups, expression of IL-4 and IL-12 was not detectable. Thus, during atherosclerosis, an oral dose of THC at 1 mg kg^{-1} seems to exert anti-inflammatory activity through suppression of the T_H1 response, resulting in a shift in the T_H1/T_H2 balance. The inhibitory effect of THC was bell-shaped, with neither the low nor the high doses having an anti-inflammatory effect.

Early in atherosclerosis development, endothelial dysfunction in

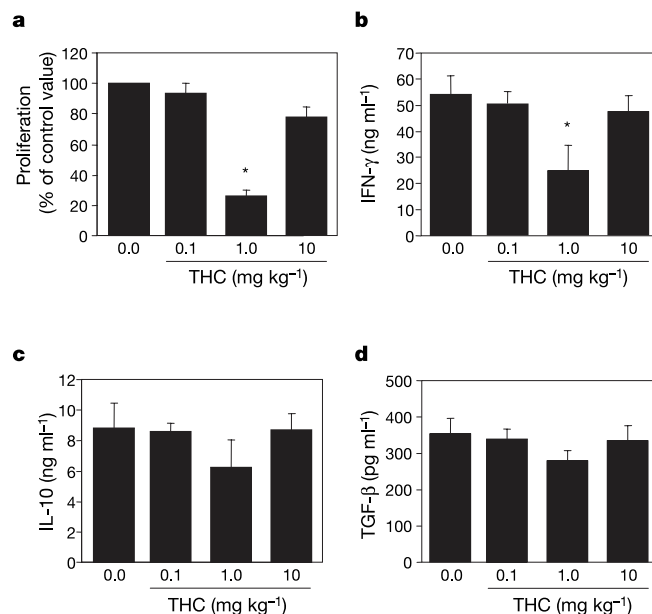


Figure 4 THC reduces proliferative responses and inhibits T_H1 polarization. **a–d**, Isolated splenocytes from THC-treated (0.1 mg kg^{-1} ; 1 mg kg^{-1} ; 10 mg kg^{-1}) or control ApoE^{-/-} mice fed the high-cholesterol diet were stimulated with Con A for 72 h (**a**) and the proliferation rate determined using the non-radioactive MTS colorimetric assay; or were stimulated for 48–72 h and concentrations of IFN- γ (**b**), IL-10 (**c**) and TGF- β (**d**) in the cell culture supernatants were determined by ELISA. Values represent mean values $\pm$ s.e.m.; $n = 6$ mice per group; single asterisk, $P < 0.05$ (compared with control).

response to cardiovascular risk factors triggers recruitment of leukocytes (monocytes/macrophages and T lymphocytes) into the vessel wall. We addressed the possibility of whether THC treatment inhibits leukocyte migration using a functional *in vitro* experiment. Thioglycollate-elicited peritoneal macrophages were isolated from ApoE^{-/-} or wild-type mice and assayed for their migration capacity in response to monocyte chemoattractant protein-1 (MCP-1). Treatment with THC at a concentration corresponding to the serum levels observed in THC-treated mice (0.6 ng ml⁻¹) significantly inhibited ApoE^{-/-} macrophage migration (Fig. 5a, b). Similar results were obtained with cells obtained from wild-type mice (data not shown). These effects were completely blocked by the CB2 antagonist SR144528 (Fig. 5a, b), and were also absent when peritoneal macrophages isolated from CB2^{-/-} mice were used (Fig. 5c, d), demonstrating that effects of THC on chemoattraction are in fact CB2 receptor-dependent. We analysed the expression levels of one of the MCP-1 receptors, CCR2, on splenocytes stimulated *in vitro* with different concentrations of

THC. As shown in Fig. 5e, THC markedly reduced CCR2 expression in splenocytes stimulated with tumour-necrosis factor- α (TNF- α). The efficacy of THC at reducing CCR2 levels was bell-shaped, with significant reduction at 0.5 and 5 ng ml⁻¹ THC, but less reduction at lower and higher doses.

We have shown that relatively low oral doses of THC (1 mg kg⁻¹), initiated after manifestation of clinically detectable artery lesions, significantly inhibit atherosclerosis progression in mice. This anti-atherosclerotic effect is probably mediated by the CB2 receptor, as it is strongly expressed by macrophages and T lymphocytes within atherosclerotic lesions; furthermore, the inhibitory effects of THC on atherosclerosis progression are abolished in the presence of a CB2 receptor antagonist. We also provide evidence that the anti-atherosclerotic properties of THC are associated with a reduction of the T_H1 response and an inhibition of monocyte/macrophage migration to the site of inflammation. Our findings *in vitro* that THC affects leukocyte migration and CCR2 expression levels at biologically relevant concentrations lend further support to our *in vivo* findings. This key influence of THC on leukocyte recruitment reinforces the potential anti-inflammatory properties of THC during atherogenesis. Our results suggest that cannabinoid derivatives with activity at the CB2 receptor may be valuable clinical targets for treating atherosclerosis. □

Methods

Reagents

Synthetic delta-9-THC Marinol (Dronabinol; Unimed Pharmaceuticals, Inc.) was dissolved at 0.1 mg ml⁻¹ in 5.5% fat milk (w/v) in water. CB2 receptor antagonist SR144528 was provided by Sanofi-Synthelabo. THC (0.1 mg kg⁻¹; 1 mg kg⁻¹ or 10 mg kg⁻¹ per day) and SR144528 (0.7 mg kg⁻¹ per day)⁸ in 1.5% fat milk (w/v) were administered orally in the drinking water. For *in vitro* experiments, delta-9-THC was purchased as a stock solution of 1 mg ml⁻¹ in methanol (Cambridge Isotope Laboratories, Inc.) and further diluted in warm medium immediately before use. All *in vitro* experiments were performed by adding corresponding dilutions of the THC vehicle (methanol) to the non-THC-treated controls.

Animals

As a model of *in vivo* atherosclerosis, we used 10-week-old male ApoE^{-/-} C57Bl/6 mice fed with a high-cholesterol diet (1.25% cholesterol, 0% cholate; Research Diets). For histological and atherosclerotic plaque development analysis, littermate ApoE^{-/-} mice were fed with a high-cholesterol diet for 4 weeks (early atherosclerotic lesion development), or 11 weeks (advanced established atherosclerosis). THC (0.1 mg kg⁻¹; 1 mg kg⁻¹ or 10 mg kg⁻¹) was administered during the last 2 weeks of the 4 week diet group ($n = 6$ per group), and during the last 6 weeks of the 11 week diet group (1 mg kg⁻¹; $n = 6$). In parallel, control ApoE^{-/-} mice (littermates) received milk without THC ($n = 6$ or 8 for the 4 and 11 week treatment groups, respectively). For CB2 receptor antagonist studies on early atherosclerotic lesion development, SR144528 was administered in parallel with THC, and control mice received corresponding dilutions of the SR vehicle (methanol; $n = 6$ per group). For cell proliferation and cytokine secretion assays, littermate ApoE^{-/-} mice were divided into four groups (control, THC 0.1 mg kg⁻¹; 1 mg kg⁻¹; 10 mg kg⁻¹; $n = 6$ per group) and fed a high-cholesterol diet for 4 weeks. THC or milk only was administered during the last 2 weeks of diet. CB2^{-/-} mice were generated as previously described³. All animal studies were approved by the local Ethical Committee.

Immunostaining

Surgical specimens of human coronary atheroma were obtained using protocols approved by the Review Committee at the University Hospital of Geneva. Immunostaining of acetone-fixed cryosections of human and mouse arteries or spleens was performed as previously described^{26,27}, using monoclonal antibodies for mouse CD4 and Mac-3 (PharMingen), and rabbit polyclonal antibody specific for both human and mouse CB2 receptor (Cayman Chemical).

Atherosclerotic lesion size and histological quantification

Atherosclerotic lesions within the thoraco-abdominal aorta and aortic roots were analysed by Sudan IV staining for lipid deposition. Quantification of lipid deposition and macrophage content (immunostaining) were performed by computer image analysis using MetaMorph6 software (Zeiss) as previously described²⁷.

Blood analysis

For measurements of cholesterol and triglyceride content, blood samples were collected at the beginning and the end of the diet treatment period. HDL and VLDL cholesterol fractions of sera were measured by fast protein liquid chromatography. THC levels in blood (after 2 and 6 weeks of treatment) were measured by gas chromatography/mass spectrometry²⁸ with a detection limit of 0.1 ng ml⁻¹.

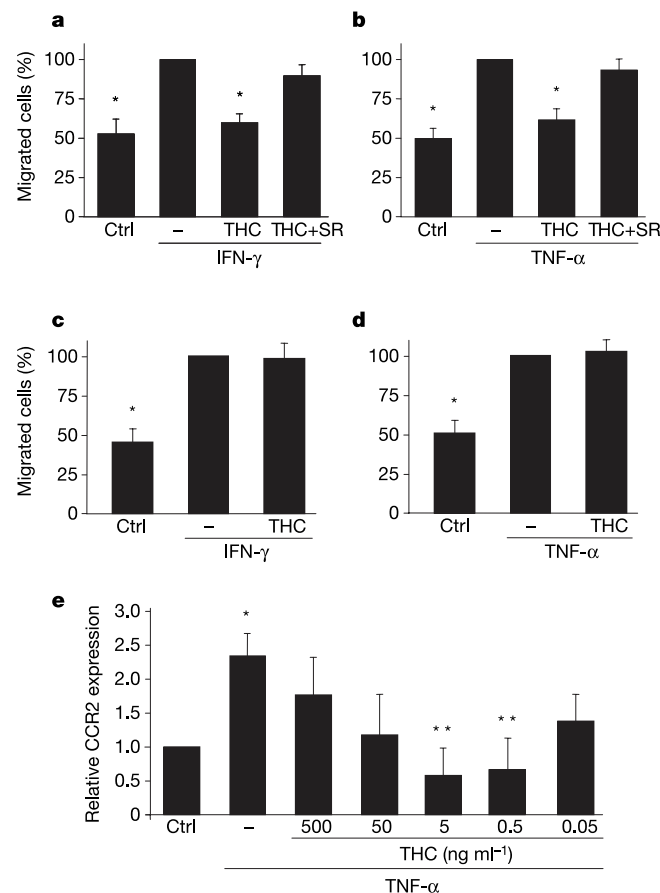


Figure 5 THC reduces migration capacity and CCR2 expression *in vitro*. **a–d**, Thioglycollate-elicited peritoneal cavity macrophages obtained from ApoE^{-/-} (**a, b**) or CB2^{-/-} (**c, d**) mice ($n = 4$ for both groups) were analysed for their *in vitro* migration response to the chemoattractant MCP-1. Cells were stimulated in duplicate with 1 ng ml⁻¹ IFN- γ (**a, c**) or 10 ng ml⁻¹ TNF- α (**b, d**) in the presence or absence of 0.6 ng ml⁻¹ THC and 1 μ M of the CB2 antagonist SR144528. Data represent mean values $\pm$ s.e.m. Single asterisk, $P < 0.05$ (compared with IFN- γ or TNF- α -stimulated cells). **e**, Isolated splenocytes obtained from ApoE^{-/-} mice ($n = 4$) were stimulated with 10 ng ml⁻¹ TNF- α in the presence or absence of different doses of THC (0.05–500 ng ml⁻¹). Relative expression levels of CCR2 messenger RNA were determined by quantitative real-time PCR. Data represent mean values $\pm$ s.e.m. Single asterisk, $P < 0.05$ (compared with control, unstimulated cells); double asterisk, $P < 0.05$ (compared with TNF- α -stimulated cells).

Intravital microscopy

Intravital microscopy of the microvasculature of the mesentery was performed as previously described²⁰. To visualize leukocytes, animals were injected intravenously with 50 µl of 0.05% rhodamine 6G (Sigma Aldrich) immediately before microscopy²⁹. ApoE^{-/-} mice were divided into three groups (control, THC; THC + SR144528; $n = 4$ per group) and fed with a high-cholesterol diet for 4 weeks. THC (1 mg kg⁻¹) and SR144528 were administered during the last 2 weeks of diet.

Proliferation assay

Splenocytes or lymph node cells were isolated from THC- or milk-treated mice ($n = 6$ per group) and cultured in 96-well plates at a concentration of 5×10^6 cells ml⁻¹. Culture medium consisted of RPMI 1640 supplemented with 25 mM HEPES buffer, 2 mM L-glutamine, 100 U ml⁻¹ penicillin, 0.1 mg ml⁻¹ streptomycin and 10% heat-inactivated FBS. Cells were stimulated in triplicates with varying concentrations of concanavalin A (Con A; Sigma). After 72 h, cell proliferation was determined using a non-radioactive MTS cell proliferation assay (Promega) according to the manufacturer's guidelines.

Real-time PCR analysis

Total RNA from mouse splenocytes was isolated with Tri-reagent (MRC, Inc), and real-time polymerase chain reaction (PCR) analysis (TaqMan, Applied Bio-systems) for CCR2 expression was performed as previously described³⁰.

Cytokine analysis

For cytokine analysis, splenocytes and lymph node cells were cultured under the same conditions as described for the proliferation assay and stimulated with 2 µg ml⁻¹ Con A. Supernatants were recovered after 48 h (for IFN-γ, IL-12 and TGF-β measurement) and 72 h (for IL-4 and IL-10 measurement). Murine IFN-γ, IL-12 (p70), IL-4, IL-10 and TGF-β were assayed by enzyme-linked immunosorbent assay (ELISA) using paired antibodies according to the manufacturer's instructions (R&D Systems).

Transmigration assay

Macrophages from the peritoneal cavity of thioglycollate-injected wild-type, ApoE^{-/-} or CB2^{-/-} mice ($n = 4$ mice per group) were isolated 4 days post-injection and stimulated (in triplicate) for 4 h with 1 ng ml⁻¹ mIFN-γ or 10 ng ml⁻¹ mTNF-α in the presence of THC. Inhibition experiments were performed by 10 min pre-incubation with 1 µM of the specific CB2 antagonist SR144528. Stimulated and unstimulated cells were transferred into the upper compartment of transwell filter migration chambers. Migration was triggered by the addition of 1 nM MCP-1 (R&D Systems) to the lower transwell chamber compartment. After 90 min incubation, the number of migrated cells was determined by counting (blinded observers) 10 microscopic fields per well.

Statistical analysis

All results are expressed as mean ± s.e.m. Differences between the values were considered significant at $P < 0.05$ using the two-tailed Student's t -test.

Received 26 October 2004; accepted 21 January 2005; doi:10.1038/nature03389.

- Libby, P. Inflammation in atherosclerosis. *Nature* **420**, 868–874 (2002).
- Klein, T. W. *et al.* The cannabinoid system and immune modulation. *J. Leukoc. Biol.* **74**, 486–496 (2003).
- Buckley, N. E. *et al.* Immunomodulation by cannabinoids is absent in mice deficient for the cannabinoid CB(2) receptor. *Eur. J. Pharmacol.* **396**, 141–149 (2000).
- Rinaldi-Carmona, M. *et al.* SR144528, the first potent and selective antagonist of the CB2 cannabinoid receptor. *J. Pharmacol. Exp. Ther.* **284**, 644–650 (1998).
- Libby, P., Ridker, P. M. & Maseri, A. Inflammation and atherosclerosis. *Circulation* **105**, 1135–1143 (2002).
- Mach, F. Statins as immunomodulatory agents. *Circulation* **109** (suppl.), II15–II17 (2004).
- Srivastava, M. D., Srivastava, B. I. & Brouhard, B. Δ9 tetrahydrocannabinol and cannabidiol alter cytokine production by human immune cells. *Immunopharmacology* **40**, 179–185 (1998).
- Zhu, L. X. *et al.* Δ9-tetrahydrocannabinol inhibits antitumor immunity by a CB2 receptor-mediated, cytokine-dependent pathway. *J. Immunol.* **165**, 373–380 (2000).
- Yuan, M. *et al.* Δ9-Tetrahydrocannabinol regulates Th1/Th2 cytokine balance in activated human T cells. *J. Neuroimmunol.* **133**, 124–131 (2002).
- Lyman, W. D., Sonett, J. R., Brosnan, C. F., Elkin, R. & Bornstein, M. B. Δ9-Tetrahydrocannabinol: a novel treatment for experimental autoimmune encephalomyelitis. *J. Neuroimmunol.* **23**, 73–81 (1989).
- Malfait, A. M. *et al.* The nonpsychoactive cannabis constituent cannabidiol is an oral anti-arthritis therapeutic in murine collagen-induced arthritis. *Proc. Natl Acad. Sci. USA* **97**, 9561–9566 (2000).
- Sulcova, E., Mechoulam, R. & Fride, E. Biphasic effects of anandamide. *Pharmacol. Biochem. Behav.* **59**, 347–352 (1998).
- Brenneisen, R., Egli, A., Elsohly, M. A., Henn, V. & Spiess, Y. The effect of orally and rectally administered Δ9-tetrahydrocannabinol on spasticity: a pilot study with 2 patients. *Int. J. Clin. Pharmacol. Ther.* **34**, 446–452 (1996).
- Chesher, G. B., Bird, K. D., Jackson, D. M., Perrignon, A. & Starmer, G. A. The effects of orally administered Δ9-tetrahydrocannabinol in man on mood and performance measures: a dose-response study. *Pharmacol. Biochem. Behav.* **35**, 861–864 (1990).
- Lichtman, A. H., Poklis, J. L., Poklis, A., Wilson, D. M. & Martin, B. R. The pharmacological activity of inhalation exposure to marijuana smoke in mice. *Drug Alcohol Depend.* **63**, 107–116 (2001).
- Varvel, S. A., Hamm, R. J., Martin, B. R. & Lichtman, A. H. Differential effects of Δ9-THC on spatial reference and working memory in mice. *Psychopharmacology (Berl.)* **157**, 142–150 (2001).
- Nakashima, Y., Plump, A. S., Raines, E. W., Breslow, J. L. & Ross, R. ApoE-deficient mice develop lesions of all phases of atherosclerosis throughout the arterial tree. *Arterioscler. Thromb.* **14**, 133–140 (1994).
- Reddick, R. L., Zhang, S. H. & Maeda, N. Atherosclerosis in mice lacking apo E. Evaluation of lesional

- development and progression. *Arterioscler. Thromb.* **14**, 141–147 (1994).
- Tangirala, R. K., Rubin, E. M. & Palinski, W. Quantitation of atherosclerosis in murine models: correlation between lesions in the aortic origin and in the entire aorta, and differences in the extent of lesions between sexes in LDL receptor-deficient and apolipoprotein E-deficient mice. *J. Lipid Res.* **36**, 2320–2328 (1995).
 - Johnson, Z. *et al.* Interference with heparin binding and oligomerization creates a novel anti-inflammatory strategy targeting the chemokine system. *J. Immunol.* **173**, 5776–5785 (2004).
 - Daugherty, A. & Rateri, D. L. T lymphocytes in atherosclerosis: the yin-yang of Th1 and Th2 influence on lesion formation. *Circ. Res.* **90**, 1039–1040 (2002).
 - Moeller, F. & Nielsen, L. B. Aortic recruitment of blood lymphocytes is most pronounced in early stages of lesion formation in apolipoprotein-E-deficient mice. *Atherosclerosis* **168**, 49–56 (2003).
 - Song, L., Leung, C. & Schindler, C. Lymphocytes are important in early atherosclerosis. *J. Clin. Invest.* **108**, 251–259 (2001).
 - Benagiano, M. *et al.* T helper type 1 lymphocytes drive inflammation in human atherosclerotic lesions. *Proc. Natl Acad. Sci. USA* **100**, 6658–6663 (2003).
 - Laurat, E. *et al.* In vivo downregulation of T helper cell 1 immune responses reduces atherogenesis in apolipoprotein E-knockout mice. *Circulation* **104**, 197–202 (2001).
 - Mulhaupt, F. *et al.* Statins (HMG-CoA reductase inhibitors) reduce CD40 expression in human vascular cells. *Cardiovasc. Res.* **59**, 755–766 (2003).
 - Kwak, B. R. *et al.* Reduced connexin43 expression inhibits atherosclerotic lesion formation in low-density lipoprotein receptor-deficient mice. *Circulation* **107**, 1033–1039 (2003).
 - Giroud, C. *et al.* Δ⁹-THC, 11-OH-Δ⁹-THC and Δ⁹-THCCOOH plasma or serum to whole blood concentrations distribution ratios in blood samples taken from living and dead people. *Forensic Sci. Int.* **123**, 159–164 (2001).
 - Hickey, M. J., Bullard, D. C., Issekutz, A. & James, W. G. Leukocyte-endothelial cell interactions are enhanced in dermal postcapillary venules of MRL/lpr (lupus-prone) mice: roles of P- and E-selectin. *J. Immunol.* **168**, 4728–4736 (2002).
 - Veillard, N. R., Steffens, S., Burger, F., Pelli, G. & Mach, F. Differential expression patterns of proinflammatory and antiinflammatory mediators during atherogenesis in mice. *Arterioscler. Thromb. Vasc. Biol.* **24**, 2339–2344 (2004).

Acknowledgements This work was supported by grants from the Swiss National Science Foundation to F.M. and J.-L.E., from the Ernst and Lucie Schmidheiny Foundation (Geneva) to F.M., and from the Foundation for Medical Research (France) to C.A. The authors of this manuscript belong to the European Vascular Genomics Network (<http://www.evgn.org>), a Network of Excellence supported by the European Community's sixth Framework Programme for Research, Priority 1. We would like to thank M. Kosco-Vilbois for helpful discussion in preparing the manuscript, and M.-L. Bochaton-Piallat (Pathology Department) for technical advice.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.M. (Francois.Mach@medecine.unige.ch).

Impairment of angiogenesis and cell migration by targeted aquaporin-1 gene disruption

Samira Saadoun*, Marios C. Papadopoulos*, Mariko Hara-Chikuma & A. S. Verkman

Departments of Medicine and Physiology, Cardiovascular Research Institute, University of California, San Francisco, San Francisco, California 94143, USA

* These authors contributed equally to this work

Aquaporin-1 (AQP1) is a water channel protein expressed widely in vascular endothelia, where it increases cell membrane water permeability^{1–3}. The role of AQP1 in endothelial cell function is unknown. Here we show remarkably impaired tumour growth in AQP1-null mice after subcutaneous or intracranial tumour cell implantation, with reduced tumour vascularity and extensive necrosis. A new mechanism for the impaired angiogenesis was established from cell culture studies. Although adhesion and proliferation were similar in primary cultures of aortic endothelia from wild-type and from AQP1-null mice, cell migration was greatly impaired in AQP1-deficient cells, with abnormal vessel formation *in vitro*. Stable transfection of non-endothelial cells with AQP1 or with a structurally different water-selective trans-

porter (AQP4) accelerated cell migration and wound healing *in vitro*. Motile AQP1-expressing cells had prominent membrane ruffles at the leading edge with polarization of AQP1 protein to lamellipodia, where rapid water fluxes occur. Our findings support a fundamental role of water channels in cell migration, which is central to diverse biological phenomena including angiogenesis, wound healing, tumour spread and organ regeneration.

AQP1 protein is expressed strongly in proliferating microvessels in human⁴ and rat⁵ malignant brain tumours, bone marrow microvessels in human multiple myeloma⁶, and proliferating microvessels in chick embryo chorioallantoic membrane⁷. Thus, indirect evidence supports a role for AQP1 in microvessel formation/function. Increased aquaporin expression has also been observed

in malignant tumour cells^{4,8}, indicating a possible role for water channels in tumour growth and spread. Here we report that targeted AQP1 gene disruption in mice reduces angiogenesis *in vivo*. By using cell culture models of AQP1 deficiency and overexpression we discovered an unexpected role for AQP1 in cell migration.

Tumour angiogenesis was studied in a well-established mouse model^{9,10}. We implanted melanoma cells subcutaneously in wild-type and AQP1-null mice. In an outbred genetic strain of mice (CD1), tumour growth was greatly slowed in the AQP1-deficient mice (Fig. 1a, left) and associated with enhanced survival (Fig. 1a, right). In control studies, tumour growth was unaffected in mice lacking AQP3 (Fig. 1a, left), a water channel that impairs urinary concentrating ability to an even greater extent than AQP1 (ref. 11). Slow growth of subcutaneous melanoma was also found in syn-

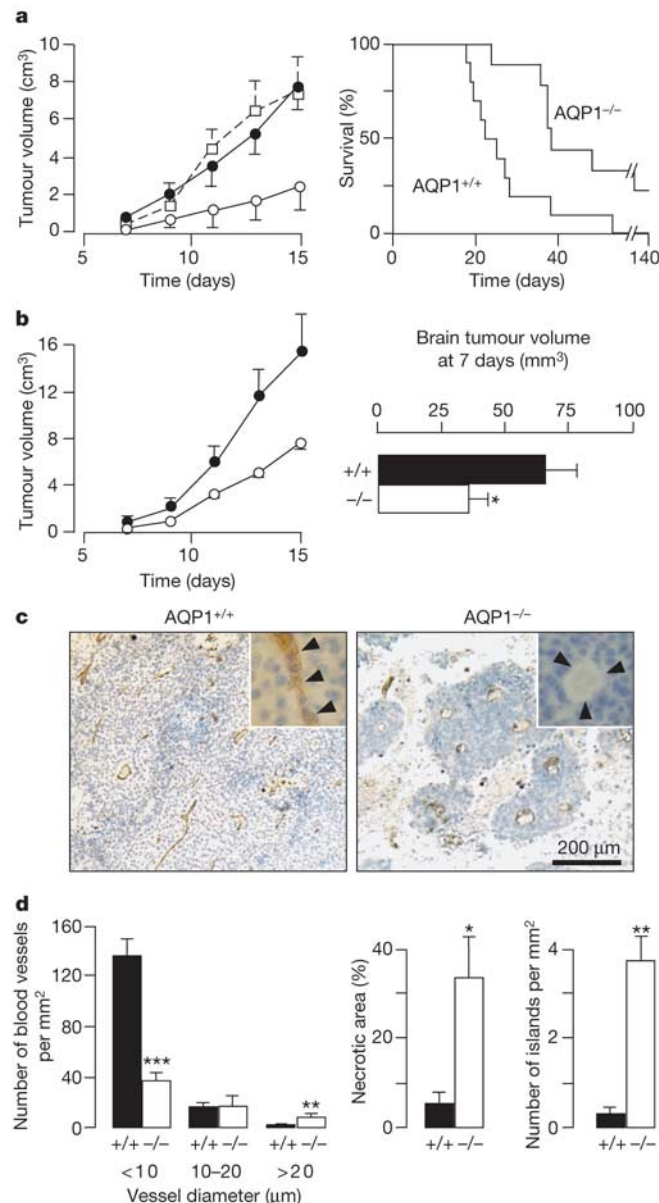


Figure 1 Reduced tumour growth in AQP1-null mice. **a**, Left, growth of subcutaneous melanoma in wild-type (AQP1^{+/+}, filled circles), AQP1 null (AQP1^{-/-}, open circles) and AQP3 null (AQP3^{-/-}, open squares) CD1 mice ($n=5-10$, $P<0.001$ comparing AQP1^{+/+} with AQP1^{-/-} or AQP3^{-/-} with AQP1^{-/-}). Right, survival of mice with subcutaneous melanoma ($n=9-10$, $P<0.001$). **b**, Left, growth of subcutaneous melanoma in C57BL/6 mice ($n=5$ each, $P<0.001$). Filled circles, AQP1^{+/+}; open

circles, AQP1^{-/-}. Right, volume of brain melanoma ($n=5-7$). **c**, Tumour stained with isolectin-B4 (brown). Insets, tumour vessels immunostained for AQP1. **d**, Left, number of vessels; centre, percentage necrotic area; right, number of islands in subcutaneous melanoma. Where error bars are shown, results are means and s.e.m.; six mice per group. Asterisk, $P<0.05$; two asterisks, $P<0.01$; three asterisks, $P<0.001$.

genic mice (C57BL/6) lacking AQP1 (Fig. 1b, left) with corresponding increased survival (not shown), and of a brain tumour as assessed histologically at 7 days after intracranial implantation of melanoma cells (Fig. 1b, right). Thus, the ability of AQP1 deficiency to limit tumour growth is independent of mouse strain, tumour location and urinary concentrating ability. A consistent histological finding in tumours of AQP1-null mice was a markedly lower density of microvessels and the presence of islands of viable tumour cells surrounded by necrotic tissue (Fig. 1c, d). All vessels in tumours of wild-type mice immunostained strongly for AQP1, with no staining seen in tumour cells or vessels of AQP1-null mice (Fig. 1c, insets). Light and electron microscopic examination of skeletal muscle and skin revealed normal vessel number and morphology in AQP1 deficiency (not shown). On the basis of these findings, we postulated that AQP1 deletion impairs tumour microvessel proliferation, which produces extensive tumour necrosis.

To confirm this hypothesis, we studied angiogenesis in a tumour-independent model involving subcutaneous implantation of reconstituted basement membrane (Matrigel) containing angiogenic factors (basic fibroblast growth factor (bFGF) or vascular endothelial growth factor (VEGF))¹². The Matrigel mass was removed after 5 days. On gross examination, little vascularity was seen in

control Matrigel specimens lacking growth factor, whereas Matrigel specimens containing bFGF or VEGF were reddish-brown with greater staining of specimens from wild-type mice (Fig. 2a, left). Histological examination revealed substantially more vessel-like structures in the growth-factor-supplemented Matrigels from wild-type mice (Fig. 2a, middle). Total Matrigel haemoglobin, a measure of intact vessel formation¹², was also increased significantly in the bFGF-containing and VEGF-containing Matrigels from wild-type mice in comparison with those from AQP1-null mice (Fig. 2a, right). The Matrigel findings confirm impaired angiogenesis in AQP1-null mice.

We postulated from the tumour and Matrigel results that an intrinsic difference in AQP1-expressing endothelial cells might account for the impaired angiogenesis in AQP1 deficiency. To test this hypothesis, intrinsic endothelial cell functions required for angiogenesis, such as proliferation, adhesion and migration, were studied with the use of cultured endothelial cells. Endothelial cells were generated from mouse aorta¹² and used 7–10 days after plating. Cells from wild-type and AQP1-null mice had similar appearance and growth (Fig. 2b). More than 90% of cells from wild-type mice stained for AQP1; there was no staining in cells from AQP1-null mice (Fig. 2c, left). Functional analysis of osmotic water per-

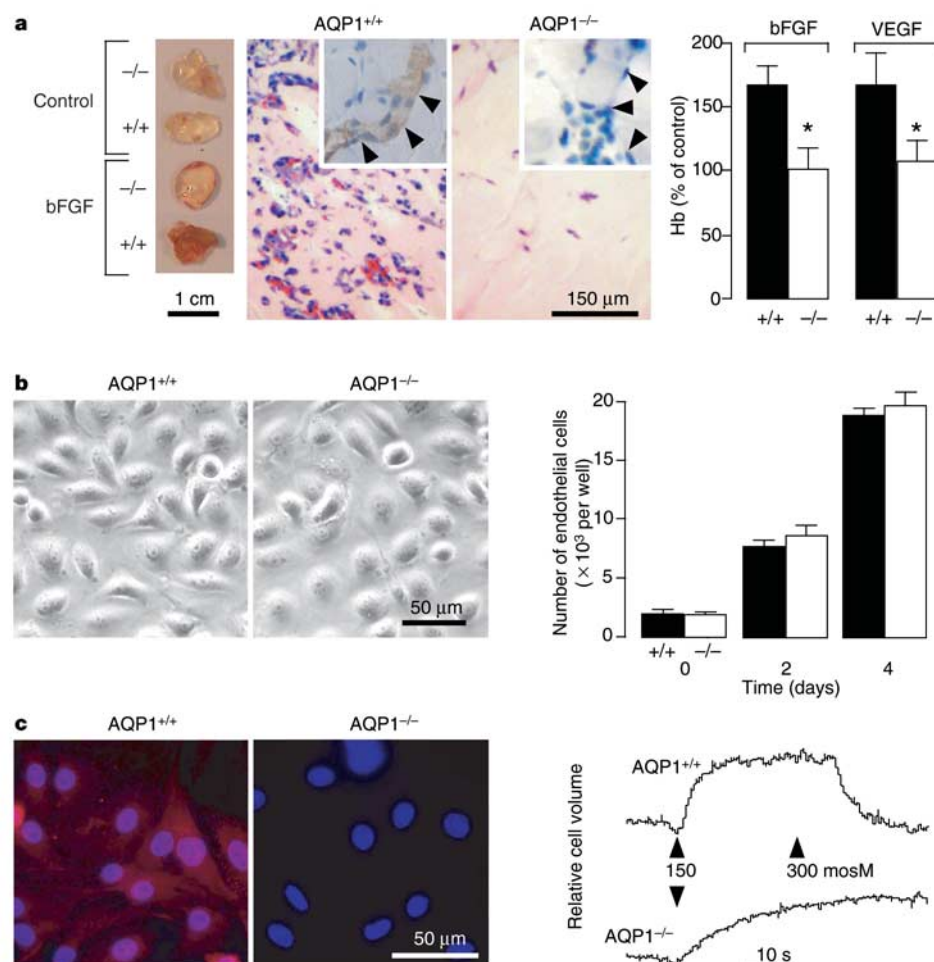


Figure 2 Angiogenesis in Matrigel and characterization of cultured endothelial cells. **a**, Left, representative Matrigel pellets; centre, increased vascularity (haematoxylin and eosin stain); right, higher haemoglobin content of growth factor-supplemented Matrigel from AQP1^{+/+} compared with AQP1^{-/-} mice (*n* = 8–9). Insets, vessels in Matrigel (arrowheads) immunostained for AQP1. **b**, Left, phase-contrast micrograph of cultured

endothelial cells. Right, proliferation of AQP1^{+/+} and AQP1^{-/-} endothelial cells (*n* = 12 each, differences not significant). **c**, Left, AQP1 immunostaining (red) with 4,6-diamidino-2-phenylindole counterstain (blue). Right, water permeability of plasma membrane, measured by calcein fluorescence quenching in response to osmotic gradients. Where error bars are shown, results are means and s.e.m. Asterisk, *P* < 0.05.

meability with the use of a calcein fluorescence quenching method¹³ revealed 2.5-fold greater water permeability in the AQP1-expressing endothelial cells (Fig. 2c, right; $n = 4$ each, $P < 0.01$).

Control and AQP1-deficient endothelial cells were used in comparative measurements of cell adhesion, migration, invasion, and cord formation, assayed in accordance with established procedures^{10,14–16}. Cell adhesion, as quantified from the number of cells adhering to a gelatin support within 4 h of plating, was not significantly altered by AQP1 expression (Fig. 3a). Cell migration towards fetal bovine serum, a potent chemotactic stimulus^{17,18}, was quantified with a modified Boyden chamber. Little migration was seen towards 1% serum but there was substantial migration for 10% serum (Fig. 3b). Migration of AQP1-deficient endothelial cells towards 10% serum was decreased significantly. Even larger differences were found in a cell 'invasion' assay in which cells migrated through a Matrigel layer followed by the porous filter. The slower migration of AQP1-deficient cells than of wild-type cells was not due to altered cell size (diameter $15.8 \pm 0.3 \mu\text{m}$ compared with $15.8 \pm 0.4 \mu\text{m}$, $n = 100$ each). In a study of cord formation, a multi-step process that includes migration, endothelial cells formed nascent cord/tube-like structures within 3 h. There was a significant

decrease in the number of small-diameter cords in AQP1-deficient cultures at 24 h and an increase in the number of large-diameter structures (Fig. 3c), similar to the findings in tumours. The total length of cords was also reduced in AQP1-null cultures ($1.9 \pm 0.2 \text{ mm mm}^{-2}$ compared with $2.9 \pm 0.1 \text{ mm mm}^{-2}$; $n = 11$ and $n = 17$, respectively; $P < 0.001$). As an independent test of cell migration, we followed wound closure after the removal of a 300–500- μm -wide strip of cells on a confluent cell monolayer¹⁹. Wound closure was significantly slowed in AQP1-deficient cells, as seen in monolayers photographed 24 h after stripping (Fig. 3d, left) and quantified as closure speed (Fig. 3d, right).

These results provide compelling evidence for the involvement of AQP1 in endothelial cell migration, offering a mechanism for impaired angiogenesis in AQP1 deficiency *in vivo*. Because mouse vascular smooth muscle cells and fibroblasts do not express AQP1 (not shown), they cannot account for the impaired angiogenesis. However, the data above do not explain how AQP1 function is linked to cell migration or whether aquaporins facilitate the migration of non-endothelial cells. Cell migration involves the transient formation of membrane protrusions (lamellipodia and cell membrane ruffles) at the leading edge of the cell; these are

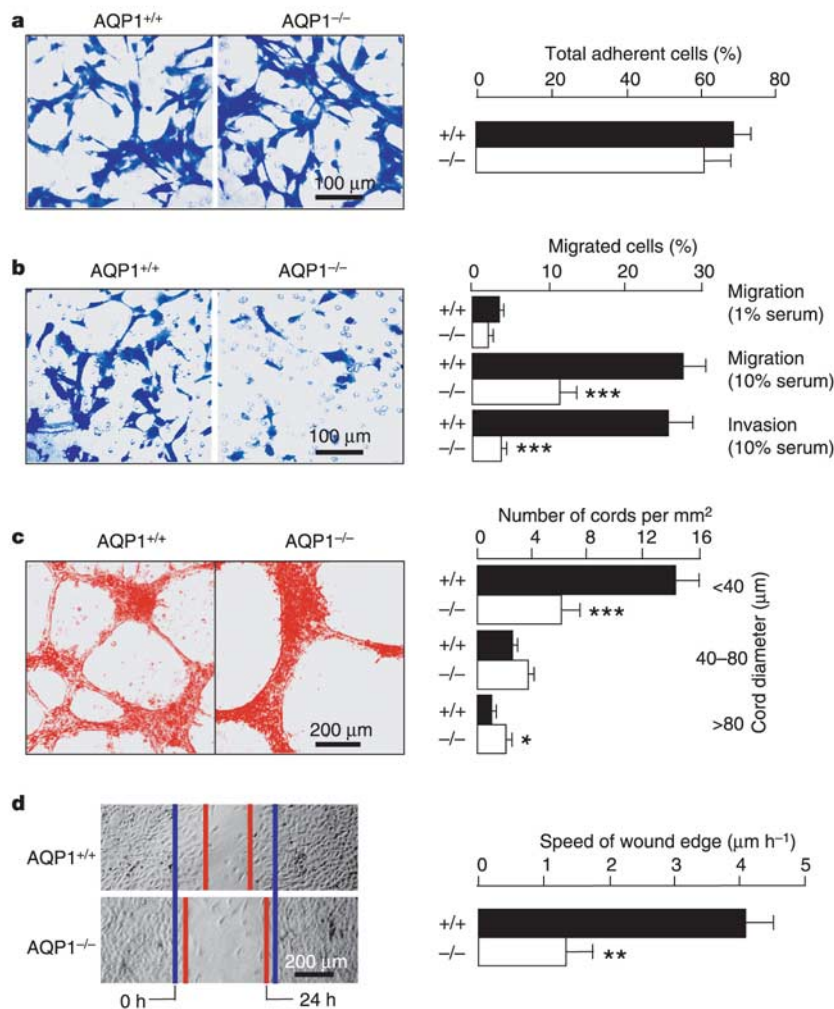


Figure 3 Impaired migration of endothelial cells lacking AQP1. **a**, Adhesion of endothelial cells (blue) on transwell filters shown (before scraping) (left) and quantified as explained in the text (right) ($n = 9$ –10 cultures, difference not significant). **b**, Migrated endothelial cells shown (after scraping) (left) with data summary (right) ($n = 16$ –20). **c**, Left, endothelial network of 'cords' (red). Right, summary of numbers of cords of indicated diameters

($n = 11$ –18). **d**, Wound healing of cultured endothelial cells (left) seen using phase contrast (blue, initial wound edge; red, after 24 h) and quantified as the wound edge speed (right) ($n = 4$ per group). Where error bars are shown, results are means and s.e.m. Asterisk, $P < 0.05$; two asterisks, $P < 0.01$; three asterisks, $P < 0.001$.

thought to require rapid local changes in ion fluxes and cell volume^{20–23}, probably accompanied by rapid transmembrane water movement²⁴. If transmembrane water movement is a fundamental determinant of cell motility, we predicted that a different water channel would also accelerate cell migration, and that water-permeability-dependent migration would be seen in many cell types. To test this prediction, wound healing and cell migration were studied in non-endothelial cells: Chinese hamster ovary (CHO) cells and Fisher rat thyroid epithelial (FRT) cells, after stable transfection with control plasmid (encoding green fluorescent protein), or with plasmids encoding AQP1, or with a structurally different water-selective transporter, AQP4 (ref. 25). Expression of AQP1 or AQP4 in plasma membrane was confirmed in the transfected cells (Fig. 4a, top), and osmotic water permeability was increased 6–8-fold compared with their respective control cells (not shown). Cell growth and adhesion were not affected by aquaporin expression, whereas cell migration through a porous

filter was remarkably enhanced in both cell types when either AQP1 or AQP4 was expressed (Fig. 4a, bottom).

Wound closure experiments supported the conclusion of increased migration in the aquaporin-expressing cells. Figure 4b (left and middle) shows accelerated closure in CHO and FRT cells expressing AQP1 or AQP4. Representative time-lapse photography showing accelerated wound closure of AQP1-expressing CHO cells is included in Supplementary Video 1. Figure 4b (right) tracks the movement over 4 h of six cells at the edge of the wound (arrow indicates the starting position) in control and AQP1-expressing CHO cells. Substantially greater directed motion was observed in each of the AQP1-expressing cells, accounting for the accelerated wound healing. An interesting observation made in about 60% of cells at the wound edge was the polarized expression of AQP1 at the leading edge of the cell membrane (Fig. 4c and Supplementary Fig. 1), where water transport is likely to occur. Polarization at the leading edge of migrating cells has been found for several transpor-

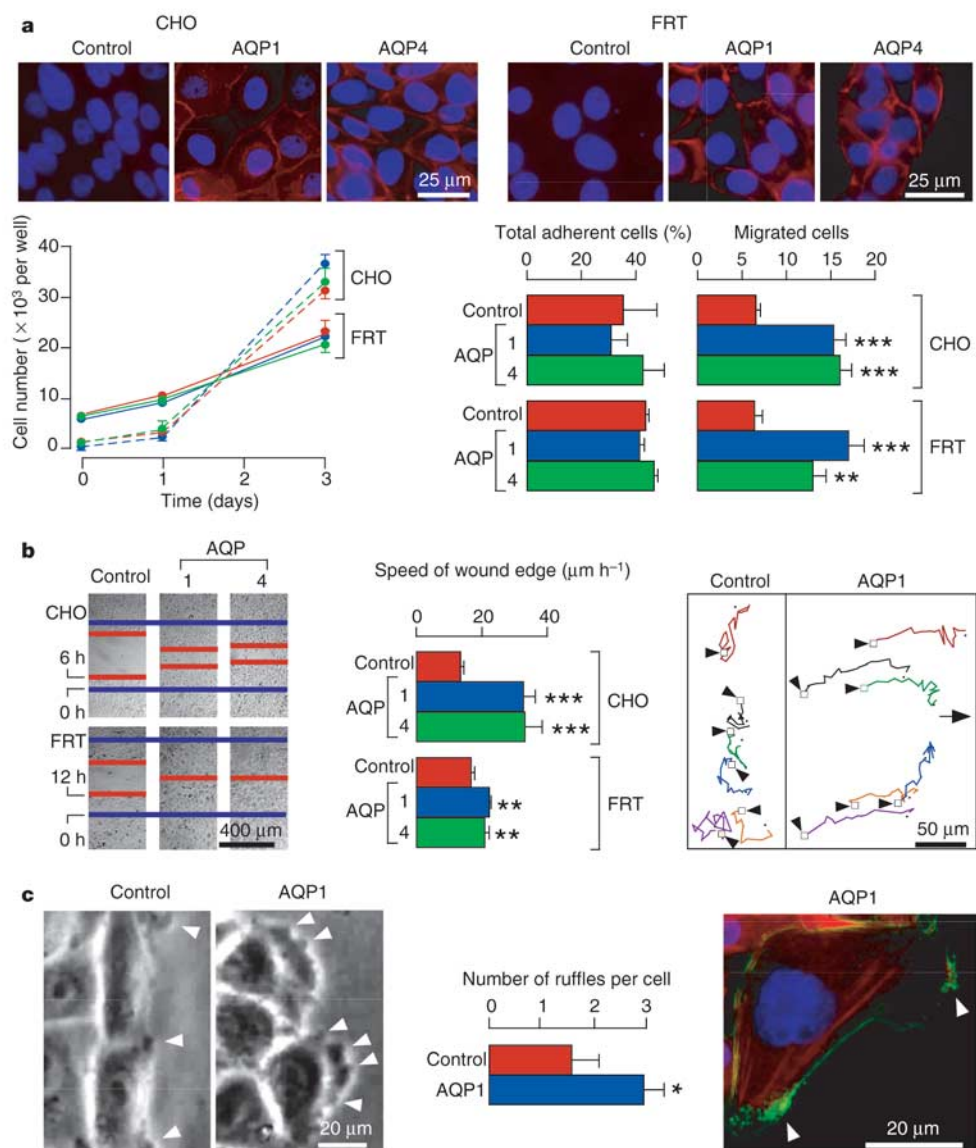


Figure 4 Increased motility after transfection with aquaporin. **a**, Top, AQP1 and AQP4 immunostaining (red) in control and transfected cells. Bottom, cell proliferation, adhesion and migration ($n=8-12$). **b**, Left, micrographs showing wound healing; centre, data summary ($n=8-12$). Right, tracks of migrating CHO cells over 4 h; arrows indicate initial positions. **c**, Left, membrane ruffles (arrowheads) in CHO cells. Centre, mean number of

ruffles (eight migrating cells per group, over 4 h). Right, AQP1 (green) polarizes to lamellipodia in migrating CHO cells. Actin is stained red. Where error bars are shown, results are means and s.e.m. Asterisk, $P < 0.05$; two asterisks, $P < 0.01$; three asterisks, $P < 0.001$.

ters involved in migration, including Na^+/H^+ and $\text{Cl}^-/\text{HCO}_3^-$ exchangers, and the $\text{Na}^+/\text{HCO}_3^-$ co-transporter²¹.

We postulated that polarization of AQP1 might alter the rate of formation of cell membrane protrusions. Time-lapse phase-contrast microscopy was performed to quantify the dynamics of cell membrane protrusions in CHO cells at the border of the wound (Supplementary Video 2). AQP1 expression produced more protrusions (Fig. 4c) and shorter mean residence time of protrusions (14 ± 1 s compared with 36 ± 4 s, eight cells analysed per group, $P < 0.001$). Together, these results indicate that aquaporins might accelerate cell migration by facilitating the rapid turnover of cell membrane protrusions at the leading edge. Cell migration and membrane ruffling occur even without aquaporins, although more slowly, because most lipid bilayers are moderately permeable to water.

It has been proposed that actin cleavage and ion uptake at the tip of a lamellipodium create local osmotic gradients that drive the influx of water across the cell membrane^{21,23,24}. Water entry then increases local hydrostatic pressure to cause cell membrane protrusion, which might create space for actin polymerization. Our findings indicate that aquaporins might provide the major pathway for water entry into lamellipodia, though alternative mechanisms might also contribute to aquaporin-dependent cell migration, such as aquaporin interactions with cytoskeletal elements or ion transporters. Aquaporins are therefore potential pharmacological targets: aquaporin inhibition in endothelial and tumour cells might limit tumour growth and spread, whereas aquaporin induction might accelerate wound healing and facilitate organ regeneration. □

Methods

Mice

AQP1-null and AQP3-null mice were generated by targeted gene disruption as described^{11,26}. Experiments were performed on weight-matched and sex-matched wild-type and AQP1-null mice in CD1 or C57BL/6 genetic backgrounds. Investigators were blinded to the genotype for all experiments. Protocols were approved by the UCSF Committee on Animal Research.

Tumour cell implantation

B16F10 melanoma cells (ATCC) were cultured in DMEM medium supplemented with 4 mM L-glutamine, 4.5 g l⁻¹ dextrose and 10% fetal bovine serum. B16F10 cells (10^6) in 200 μ l PBS were injected subcutaneously between the shoulder blades. Tumour length (L) and width (W) were measured with a caliper for the estimation of tumour volume as W^2L (ref. 9). In some mice, 5 μ l PBS containing 10^5 B16F10 cells was stereotactically injected into the right cerebral hemisphere as described²⁷. Brain tumour volume was measured 7 days after implantation by summing the tumour areas of 1-mm-thick coronal brain slices.

In vivo Matrigel assay

Each anaesthetized mouse received two 0.5-ml injections of Matrigel (without or with 20 nM bFGF plus 25 mg heparin or 100 ng ml⁻¹ VEGF (BD Biosciences)) under abdominal skin¹². Matrigel pellets were harvested after 5 days, digested with dispase (Sigma-Aldrich) and haemoglobin content was determined by Drabkin's method (Ricca Chemical Company). Matrigel pellets were examined histologically after staining with haematoxylin and eosin.

Histology

Tumours were fixed in formalin and embedded in paraffin. Sections were immunostained with a polyclonal rabbit anti-AQP1 antibody²⁶ followed by biotinylated goat anti-rabbit antibody or biotinylated isoelectin-B4 (Vector Labs). Isoelectin-B4 selectively stains mouse blood vessels²⁸. Sections were then treated with avidin-horseradish peroxidase and diaminobenzidine, incubated with 3% H₂O₂ to bleach melanin, and counterstained with haematoxylin.

Endothelial cell culture

Mouse aortic endothelial cells were isolated with collagenase type 2 (Worthington Biochem) as described¹² and cultured on fibronectin in endothelial serum-free medium (Gibco) supplemented with 20 ng ml⁻¹ bFGF and 10 ng ml⁻¹ epidermal growth factor. More than 90% of cells were of endothelial origin as assessed by von Willebrand factor immunostaining (Dako) and diI-Ac-LDL uptake (Biomedical Technologies).

Water permeability measurements

Cells grown on coverslips were loaded with calcein by incubation for 30 min with 10 μ M calcein acetoxymethyl ester (Molecular Probes) and mounted in a perfusion chamber designed for rapid solution exchange. Solutions were exchanged from 300 to 150 mosM PBS and the rate of change of calcein fluorescence was monitored as described¹³.

Adhesion and proliferation

Medium was exchanged 4 h after plating. Adhesion was defined as the percentage of plated cells remaining immediately after medium exchange. For proliferation, the number of cells was estimated on days 1–4. Cell number was determined with a chromogenic assay kit in 96-well plates (Promega)¹⁴. Results were confirmed independently by cell counting.

Migration and invasion

Assays were performed with a modified Boyden chamber (Corning Costar) containing a gelatin-coated polycarbonate membrane filter (6.5 mm diameter, 8 μ m pores)^{10,16}. To study invasion, the upper surface of the filter was also coated with 20 μ l Matrigel (0.3 mg ml⁻¹). The upper chamber contained cells in DMEM plus 1% fetal bovine serum, and the lower chamber contained DMEM plus 10% fetal bovine serum (chemoattractant) or 1% fetal bovine serum (control). Cells were incubated for 6 h at 37 °C in 5% CO₂. Non-migrated cells were scraped from the upper surface of the membrane with a cotton swab. Migrated cells remaining on the bottom surface were counted after staining with Coomassie blue.

Cell volume

Cell diameter was measured after trypsinizing cells, suspending them in medium and photographing them at high magnification.

Cord formation

In vitro angiogenesis was assayed in Matrigel-coated wells as described^{10,16}. Cord formation was quantified at 24 h.

In vitro wound healing

Cells were cultured as confluent monolayers, synchronized in 1% fetal bovine serum for 24 h, and wounded by removing a 300–500- μ m strip of cells across the well with a standard 200- μ l pipette tip¹⁹. The wounded monolayers were washed twice to remove non-adherent cells. Wound healing was quantified as the average linear speed of the wound edges over 24 h. In some experiments, time-lapse photography of the wound edges was performed (121 frames over 2 min, or 25 frames over 4 h).

Cell lines

FRT and CHO-K1 cells were stably transfected with plasmids to express green fluorescent protein (control), AQP1 or AQP4, and used in adhesion, proliferation, migration and wound-healing assays²⁵. More than 95% of cells expressed the respective proteins.

Statistical analysis

Statistical analysis was performed with a two-tailed indirect Student's *t*-test, or analysis of variance with a post-hoc Student–Newman–Keuls and log-rank test. WinStat (RF Software) and XLStat (AddinSoft) software were used.

Received 2 September 2004; accepted 7 February 2005; doi:10.1038/nature03460.

- Nielsen, S., Smith, B. L., Christensen, E. I. & Agre, P. Distribution of the aquaporin CHIP in secretory and absorptive epithelia and capillary endothelia. *Proc. Natl Acad. Sci. USA* **90**, 7275–7279 (1993).
- Hasegawa, H., Lian, S. C., Finkbeiner, W. E. & Verkman, A. S. Extrarenal tissue distribution of CHIP28 water channels by *in situ* hybridization and antibody staining. *Am. J. Physiol.* **266**, C893–C903 (1994).
- Carter, E. P., Olvezky, B. P., Matthay, M. A. & Verkman, A. S. High microvascular endothelial water permeability in mouse lung measured by a pleural surface fluorescence method. *Biophys. J.* **74**, 2121–2128 (1998).
- Saadoun, S., Papadopoulos, M. C., Davies, D. C., Bell, B. A. & Krishna, S. Increased aquaporin 1 water channel expression in human brain tumours. *Br. J. Cancer* **87**, 621–623 (2002).
- Endo, M., Jain, R. K., Witwer, B. & Brown, D. Water channel (aquaporin 1) expression and distribution in mammary carcinomas and glioblastomas. *Microvasc. Res.* **58**, 89–98 (1999).
- Vacca, A. *et al.* Microvessel overexpression of aquaporin 1 parallels bone marrow angiogenesis in patients with active multiple myeloma. *Br. J. Haematol.* **113**, 415–421 (2001).
- Ribatti, D. *et al.* Aquaporin-1 expression in the chick embryo chorioallantoic membrane. *Anat. Rec.* **268**, 85–89 (2002).
- Moon, C. *et al.* Involvement of aquaporins in colorectal carcinogenesis. *Oncogene* **22**, 6699–6703 (2003).
- Egami, K. *et al.* Role of host angiotensin II type 1 receptor in tumor angiogenesis and growth. *J. Clin. Invest.* **112**, 67–75 (2003).
- Miao, W. M. *et al.* Thrombospondin-1 type 1 repeat recombinant proteins inhibit tumor growth through transforming growth factor-beta-dependent and -independent mechanisms. *Cancer Res.* **61**, 7830–7839 (2001).
- Ma, T. *et al.* Nephrogenic diabetes insipidus in mice lacking aquaporin-3 water channels. *Proc. Natl Acad. Sci. USA* **97**, 4386–4391 (2000).
- Yao, L. *et al.* Contribution of natural killer cells to inhibition of angiogenesis by interleukin-12. *Blood* **93**, 1612–1621 (1999).
- Solenov, E., Watanabe, H., Manley, G. T. & Verkman, A. S. Sevenfold-reduced osmotic water permeability in primary astrocyte cultures from AQP-4-deficient mice, measured by a fluorescence quenching method. *Am. J. Physiol. Cell Physiol.* **286**, C426–C432 (2004).
- Steinle, J. J. *et al.* Eph B4 receptor signaling mediates endothelial cell migration and proliferation via the phosphatidylinositol 3-kinase pathway. *J. Biol. Chem.* **277**, 43830–43835 (2002).
- Shi, G. P. *et al.* Deficiency of the cysteine protease cathepsin S impairs microvessel growth. *Circ. Res.* **92**, 493–500 (2003).
- Trojanovskiy, B., Levchenko, T., Mansson, G., Matvijenko, O. & Holmgren, L. Angiomotin: an angiostatin binding protein that regulates endothelial cell migration and tube formation. *J. Cell Biol.* **152**, 1247–1254 (2001).
- Ishida, T. *et al.* Targeted disruption of endothelial cell-selective adhesion molecule inhibits angiogenic processes *in vitro* and *in vivo*. *J. Biol. Chem.* **278**, 34598–34604 (2003).
- Orr, A. W., Elzie, C. A., Kucik, D. F. & Murphy-Ullrich, J. E. Thrombospondin signaling through the

- calreticulin/LDL receptor-related protein co-complex stimulates random and directed cell migration. *J. Cell Sci.* **116**, 2917–2927 (2003).
19. Lee, H., Goetzl, E. J. & An, S. Lysophosphatidic acid and sphingosine 1-phosphate stimulate endothelial cell wound healing. *Am. J. Physiol. Cell Physiol.* **278**, C612–C618 (2000).
 20. Schwab, A., Schuricht, B., Seeger, P., Reinhardt, J. & Dartsch, P. C. Migration of transformed renal epithelial cells is regulated by K⁺ channel modulation of actin cytoskeleton and cell volume. *Pflügers Arch.* **438**, 330–337 (1999).
 21. Schwab, A. Function and spatial distribution of ion channels and transporters in cell migration. *Am. J. Physiol. Renal Physiol.* **280**, F739–F747 (2001).
 22. Rosengren, S., Henson, P. M. & Worthen, G. S. Migration-associated volume changes in neutrophils facilitate the migratory process *in vitro*. *Am. J. Physiol.* **267**, C1623–C1632 (1994).
 23. Lauffenburger, D. A. & Horwitz, A. E. Cell migration: a physically integrated molecular process. *Cell* **84**, 359–369 (1996).
 24. Condeelis, J. Life at the leading edge: the formation of cell protrusions. *Annu. Rev. Cell Biol.* **9**, 411–444 (1993).
 25. Yang, B., Brown, D. & Verkman, A. S. The mercurial insensitive water channel (AQP-4) forms orthogonal arrays in stably transfected Chinese hamster ovary cells. *J. Biol. Chem.* **271**, 4577–4580 (1996).
 26. Ma, T. *et al.* Severely impaired urinary concentrating ability in transgenic mice lacking aquaporin-1 water channels. *J. Biol. Chem.* **273**, 4296–4299 (1998).
 27. Papadopoulos, M. C., Manley, G. T., Krishna, S. & Verkman, A. S. Aquaporin-4 facilitates reabsorption of excess fluid in vasogenic brain edema. *FASEB J.* **18**, 1291–1293 (2004).
 28. Kawamoto, A. *et al.* Intramyocardial transplantation of autologous endothelial progenitor cells for therapeutic neovascularization of myocardial ischemia. *Circulation* **107**, 461–468 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was funded by grants from the National Institutes of Health (to A.S.V.) and by a Wellcome Trust Clinician Scientist Fellowship (to M.C.P., sponsored by S. Krishna).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.S.V. (verkman@itsa.ucsf.edu).

erratum

Mesozoic Alpine facies deposition as a result of past latitudinal plate motion

Giovanni Muttoni, Elisabetta Erba, Dennis V. Kent & Valerian Bachtadse

Nature **434**, 59–63 (2005).

In Fig. 2 of this Letter, several labels were missing. In the timescale columns on the left, the marine magnetic anomalies were incomplete (they should read, from the bottom up: Atlantic closed, BSMA, M37, M22, M17, M16, M15, M11, M3, M1, M0r, “Cretaceous Quiet Zone”), as were the stages (they should read, from the bottom up: RHAETIAN, HETTANGIAN, SINEMURIAN, PLIENSBACHIAN, TOARCICAN, AALENIAN, BAJOCIAN, BATHONIAN, CALLOVIAN, OXFORDIAN, KIMMERIDGIAN, TITHONIAN...). For the positions of the missing labels, please see the HTML version of the original paper on *Nature*’s website (www.nature.com/nature/journal/v434/n7029/full/nature03378_fs.html), where Fig. 2 has been corrected. □

addendum

Strong polarization enhancement in asymmetric three-component ferroelectric superlattices

Ho Nyung Lee, Hans M. Christen, Matthew F. Chisholm, Christopher M. Rouleau & Douglas H. Lowndes

Nature **433**, 395–399 (2005).

It has been drawn to our attention that in our Letter we may have inadvertently underplayed the significance of an earlier experimental study (ref. 11) through an unfortunate choice of wording. Our intention was not to claim the first experimental realization of asymmetric properties in a three-component superlattice (ref. 11 claimed a similar realization), but the first experimental observation of strain-enhanced ferroelectric polarization in such structures. □

Pure but not simple

Protein purification is stepping into the limelight, as proteomics researchers demand faster ways to purify more proteins. Tim Chapman looks at what is around to help them.

Protein isolation is one of the oldest 'biotechnologies', but the demands from proteomics for the purification of potentially vast numbers of proteins is driving new developments in long-established techniques. "Researchers working in genomics are looking for the next step to fully understand the results of their sequence analysis — the proteins that the genome expresses — in the context of systems biology," says Anke Cassing, associate director for corporate strategy at Qiagen in Hilden, Germany. "The delicate interplay of proteins is, of course, also of extreme interest to pharmaceutical companies, which are always on the lookout for new drug targets."

There is increasing demand from researchers producing biopharmaceuticals — antibodies and proteins used as drugs. "There's a driving force towards protein purification, separation and analysis," says Carsten Buhlmann, product manager at Agilent, based in Palo Alto, California. "Especially for the biopharmaceuticals, there's a high demand for purity of these proteins that are used for drugs and have to get

through all the regulations."

For virtually all applications, researchers need to maintain a protein's biological activity, which can rule out some purification processes. Proteins can be fragile and easily denatured, and many of the most important are insoluble in the most common media.

"If you look at the average protein, it's quite complex, it's a buzzing, vibrating molecule — it's not a fixed structure," says Allan Simpson, vice-president for product development at the protein separations division of GE Healthcare Biosciences in Uppsala, Sweden. "They're very hard to handle, they're difficult to purify, they can aggregate easily — these are very hard things to manipulate."

Proteomics workhorse

With proteins taking centre stage in many laboratories, equipment developers are rolling out a new generation of automated



Anke Cassing: systems biology is the next step.

systems to take the grind out of separating and purifying proteins of interest. Chromatographic separation is one of the basic protein-purification techniques and one platform that is emerging as a workhorse of the large proteomics lab is ÄKTAexpress. Made by GE Healthcare, this is a dedicated high-throughput multistep chromatography system for purifying histidine (His)- and glutathione-S-transferase (GST)-tagged recombinant proteins.

GE began developing the platform in the late 1990s after realizing that there were not enough trained chromatographers to produce proteins in the quantities and varieties demanded by post-genomic researchers. "We decided to see if we could automate a system that would be better than the current technologies at solving that problem," Simpson says. "Instead of taking a robot and automating the current system, we set out to develop a

QIAGEN

SMALL-SCALE SEPARATION

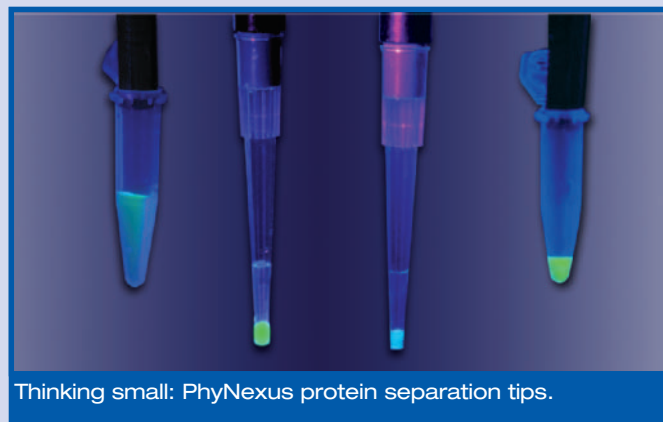
Sometimes you can do more with less. Proprietary pipette tips developed by PhyNexus in San Jose, California, promise high performance in tiny volumes with minimum fuss. The key lies in encapsulating very small quantities — just 5–10 microlitres — of protein separation resin between hydrophilic screens in the very end of the pipette tip. "The whole sample is obliged to make highly intimate contact with that resin," says Chris Hanna, vice-president of business development. "That results in high trapping efficiency for the sample. We can get a 10–20-fold increase in the target protein

sample from just a few hundred microlitres of sample, and get purities that are often over 95% with a single separation step."

As well as the technical collaboration with Caliper of Hopkinton, Massachusetts, PhyNexus has designed its PhyTips to be compatible with liquid-handling robots from Tecan of Männedorf, Switzerland, Beckman Coulter of Fullerton, California, and PerkinElmer of Boston, Massachusetts. The firm has also agreed licences to use some of the most advanced resins in its tips, including Qiagen's Ni-NTA resin for purifying histidine-tagged proteins. The combination of tips, resins and platforms makes for exquisite control of the separation process, Hanna says. "You can control the number of cycles that go back and forth through the bed, the rate they do so, the composition of the washes. We can really make that microvolume of material dance and perform at its best," he says. "To be able to do that and maintain fully functional proteins at these very small scales gives you ÄKTA-type purification off a very small amount of starting sample. People can avoid scaling up."

The tips are initially being deployed for rapid purification and enrichment of antibodies from small cultures of *Escherichia coli* for use in high-throughput cell-based assays, giving significant savings in time and money. "People are wanting to get real biological information much earlier in their screening process, and to do that they have to have the stuff properly prepped," notes Hanna. Interest is also coming from protein engineering and biopharmaceutical companies looking to miniaturize their protein-expression systems.

T.C.



Thinking small: PhyNexus protein separation tips.

new one. A robot represents nothing more than a mechanical technician — it reproduces all the successes but also all the errors, so you don't move forward in your science."

The company produced a high-throughput system that needs no specialized knowledge to operate and can be programmed to carry out up to four common purification steps starting with affinity purification. "The skill of chromatography is sitting within the system," says Simpson. The software is written as a series of wizards, each representing a single step. GE will shortly be launching a software package for the purification of monoclonal antibodies.

The basic four-module set-up can purify up to 2,500 proteins a year, each module producing up to 50 mg of protein per run. A twin-pack version is aimed at smaller labs wanting to purify up to 1,000 proteins a year.

The new protein-purification facility at Monash University in Melbourne, Australia, is deploying a 12-module ÄKTAExpress set-up for its ambitious development programme. "I'm a structural biologist, so my interest is in producing large amounts of recombinant protein for structural and functional studies," says James Whisstock, scientific director of the facility. "What's really important is that the cost of equipment is within reach of a normal university laboratory set-up. There are some very big structural biology institutes with between US\$50 million and \$100 million's worth of industrial-scale protein preparation equipment. From our point of view that's not

achievable, but we're bringing in this technology, which is going to make a huge difference to our research."

The ability to deal with many more proteins simultaneously will allow the lab to approach problems differently. "If you have a very challenging protein target and want to try 50 different constructs, at the moment it's really not feasible to do that manually one after the other," Whisstock points out. "Now, you can try the same molecule from 50 different species. With parallel advances in expression technology, the whole process is simplified and really streamlined, and provides the capacity to perform that experiment. You're trying so many different things simultaneously, you're likely to get a result."

Automating innovation

Several companies are developing equipment and product ranges that can be used at different points in the proteomics pipeline, from raw cell extracts to mass spectrometry and beyond. Beckman Coulter in Fullerton, California, is rolling out its ProteomeLab family to help with everything from initial purification of cell extracts, through protein fractionation and characterization, to the ultimate steps of disease diagnosis.

"We try to link technologies together to simplify the job for what takes place at the end, which is typically mass spectrometry," says John Hobbs, group product manager for ProteomeLab. "To get to that point, a lot of people have realized that it's garbage in, garbage out. If you put crap into a mass spectrometer,



Scaling-up: ÄKTAExpress at Monash University's protein purification facility.

the results you get out will be the same."

The ProteomeLab PF 2D Protein Fractionation System automates two-dimensional chromatographic fractionation, resolving proteins by isoelectric point and hydrophobicity. The emphasis is on standardizing the protocols and techniques used by researchers. "Biologists want to be able to look at their results and see if they relate to someone else's,"

ATTRACTING ATTENTION

Magnetic beads have been used for protein separation since the 1980s, but the technology is now being adapted for new proteomic applications and use with automated platforms. The market leader in paramagnetic beads is Dynal Biotech based in Oslo, Norway, and recently acquired by life-sciences giant Invitrogen in Carlsbad, California.

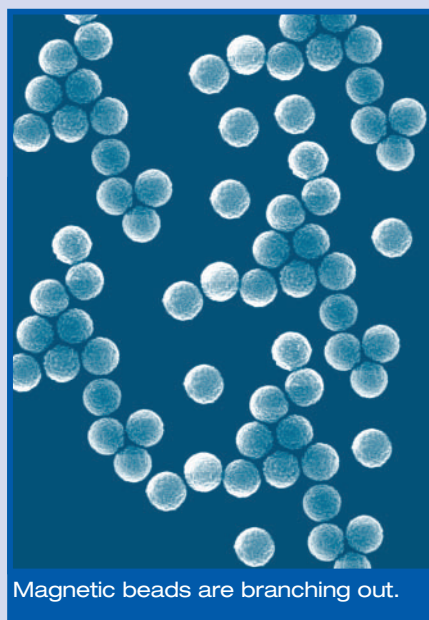
Dynal has just signed a co-marketing agreement for its Dynabead kits and Tecan's Freedom EVO automated platform, and has developed protocols for other platforms such as Beckman Coulter's Biomek FX and the KingFisher magnetic separation platform from Thermo Electron of Waltham, Massachusetts. "The main purpose of having magnetic beads is that you can automate the whole process," says Lars Korsnes, director of research and development at Dynal. "The bead technology has some advantages compared with standard chromatography systems — while it's not so easy to put a whole-blood sample into a chromatography column, with magnetic beads you can put the whole sample in."

Dynabeads, like those from some other

suppliers, are superparamagnetic, with no residual magnetism outside an applied magnetic field. They are also uniform in size, shape and surface properties. This all helps to prevent the beads clogging up an automated device, Korsnes notes.

Bead technology is also more scaleable than chromatography columns, although Dynal is concentrating on more analytical or small-scale protein isolation and protein fractionation for different applications in proteomics. The firm is currently launching a new range of beads with functionalities such as ion-exchange groups, reverse-phase chromatography and hydrophobic chemistries. New owner Invitrogen plans to apply Dynal's surface technologies to a wider range of products.

The beads are also showing promise in the challenging separation of membrane proteins. "There have been some publications where one can bind membrane proteins either before or after lysing the cells," Korsnes says. "Whether we will develop a special protocol is a question for the future, but the technology is there already."



Magnetic beads are branching out.

Hobbs notes. "As well as providing the instrument, we provide the methodology and buffers, but we do specify you have to use that method to get the full support. It's a little bit different from the normal research instrument approach, but we thought there was a need for that and it seems to be accepted."

Beckman is also currently commercializing an innovative system of protein partitioning using affinity fractionation to decrease the unwanted complexity of protein mixtures before analysis. The aim is to remove not just very abundant proteins, for example serum albumin from blood plasma, but also other proteins that are already well characterized. The firm claims that up to 95% of the proteins in a cell lysate can be removed before the full fractionation stage with less risk than other purification procedures of losing the proteins you're interested in. "There's a growing interest in what might be going away with these large-abundance proteins — albumin is a binding protein, and possibly some interesting proteins go with it," Hobbs notes.

Several big equipment producers have teamed up with smaller specialist firms to include cutting-edge reagents or media in application kits for their automated systems. Tecan in Männedorf, Switzerland, recently signed a licensing agreement to deploy the paramagnetic beads developed by Dynal Biotech in Oslo, Norway, on its Freedom EVO liquid-handling platform (see 'Attracting attention', page 796). The Robopop protein purification kits from Novagen in Madison, Wisconsin, can also be used on

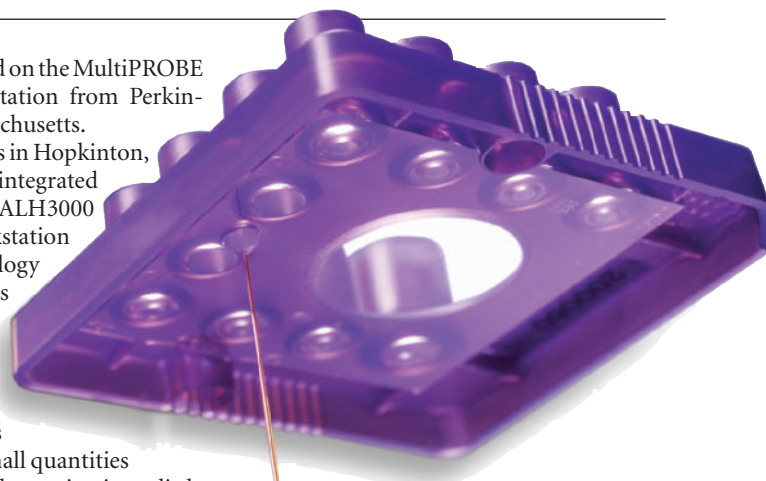
Tecan's workstation and on the MultiPROBE liquid-handling workstation from Perkin-Elmer in Boston, Massachusetts.

Caliper Life Sciences in Hopkinton, Massachusetts, has integrated into its Sciclone ALH3000 liquid-handling workstation a new column technology developed by PhyNexus based in San Jose, California (see 'Small-scale separation', page 795). The combination allows researchers to purify and enrich small quantities of up to 96 engineered proteins in as little as 15 minutes.

The market for protein purification systems has changed in the past six months, notes Mark Roskey, vice-president of marketing at Caliper, with more groups getting involved in larger-scale protein purification. "It's not at the industrial scale, but regular pharma and biotech R&D people are now trying to purify proteins in a more parallel situation. A lot of this stems from having all the genes and working with huge numbers of them to develop new drugs," he says.

Analysis and optimization

To maintain the benefits of high-throughput separation and purification, the proteins of interest must be able to pass smoothly into the next stage of the process. "Once you've got a relatively pure protein you need to determine whether it is pure, so



The LabChip90 from Caliper Life Sciences.

there's issues with analysis as well," says Roskey. A common analysis method is SDS-polyacrylamide gel electrophoresis "but we feel that that is a bottleneck," Roskey adds.

Although the established protocols of macroscale gel electrophoresis are being successfully automated (see 'Automation in two dimensions', below), many users are turning instead to microfluidic and lab-on-a-chip solutions. In January, Caliper launched the Protein Express Assay for its LabChip 90 automated electrophoresis system.

"It's a microfluidic replacement for SDS-PAGE that automates the whole process," Roskey says. "Rather than putting samples on a gel, you get them off a multiwell plate, and it does integrated separation,

AUTOMATION IN TWO DIMENSIONS

Microfluidic systems have taken over from two-dimensional (2D) gel electrophoresis techniques in some areas of protein separation and analysis, but the established methods are far from dead. Even though they can be slower and messier, the tried and tested 2D protocols still offer some advantages, especially if automation can take out most of the hassle.

"Some people will be claiming otherwise, but I think 2D still has the best levels of sensitivity when you're looking at complex samples," says Paul Orange, senior product manager at NextGen Sciences in Cambridge, UK. "With 2D you're seeing resolution of 2,000–3,000 spots on a gel — that's quite a lot of information, but people understand it. Also 2D is a very accessible technology — you can go and buy equipment relatively cheaply and get started if you're looking for a proteomics approach."

In 2003, NextGen launched the first fully automated 2D electrophoresis platform, called a2DE. The firm has now brought out a spin-off system called the a2DEoptimizer which can improve 2D separations on a variety of commercial platforms. "We've not tried to reinvent the wheel," says Orange. "We know what people are using, and know that they have lots of data and experience,

but we can help them out by automating key aspects of the process."

The heart of the a2DEoptimizer is automated gel casting, allowing researchers to create customized gradient gel profiles with a minimum of fuss. "Everyone knows casting a gradient gel gives you far superior spot resolution, separation and definition, but the problem is that these gradients can be quite tricky to pour, especially the more exotic ones," Orange says. "There's a very small number of people who can get good reproducible gels when they're pouring gradients. As we see 2D going forward, people are dealing with very small amounts of sample. They've got one or two gels they can run so they have to get the best data they can out of there."

The system also has integrated power packs that can focus samples at high voltages and reduce the time for a separation run. And it has real-time monitoring of the electrical profile. "That's of particular importance, because when you're dealing with new samples, it's very important to look at the electrical profile and tell whether you've got some kind of issue with salt content or protein content," Orange notes. "What we're doing is enabling people to get better data out, and also analyse what's going on in their system." **T.C.**



2D gel electrophoresis gets a facelift.

detection and analysis. It can do an individual protein in around 30 seconds, so you can do 96 proteins in an hour with the same quality as SDS — in fact better, because the data are digital.”

Automated electrophoresis using chips developed with Caliper is the basis of two new systems — the bench-top Experion from Bio-Rad in Hercules, California, which analyses a single Pro260 chip carrying a maximum of 10 samples in 30 minutes for the medium-scale user, and the ultra-high-throughput 5100 Automated Lab-on-a-Chip system from Agilent in Palo Alto, California. The latter is aimed at pharmaceutical companies and other laboratories needing to separate, purify and analyse thousands of proteins a day. The fully automated 5100ALP using the Protein 200 Plus LabChip kit can take up to twelve 96- or 384-well plates for overnight analysis, with each chip capable of up to 6,000 sample runs. “The real innovation with the 5100 is that you have a complete unattended solution ending up with digital data,” says Carsten Buhlmann, product manager for Agilent’s microfluidics group.

The 5100ALP has been deployed at the centralized protein-production facility for AstraZeneca in Alderley Park, UK. “We’re developing a high-throughput protein-production platform and wanted a quantitative and qualitative data-analysis method for that process,” says Paul Hawtin, senior research chemist at AstraZeneca’s UK protein group. “Previously we used SDS-PAGE gels and although they get you the information you need, they’re very cumbersome, especially when you’re using the numbers that we’re using.”

With high-throughput protein production, an integrated analysis capability is invaluable. “We were in the situation where we could do high-throughput molecular biology, high-throughput expression and also high-throughput production — once you’ve gone through that process and you’re testing a number of variables, the numbers you’ve got coming out the back end are incredibly large,”



Allan Simpson: membrane proteins will be “the pot of gold”.

Hawtin notes. The AstraZeneca team also uses the ÄKTApurify to follow up with larger-scale protein production.

Kits and columns

Researchers who don’t need automation can choose from an increasing selection of specialized purification and fractionation kits. Kits for affinity purification of recombinant proteins tagged with His₆, GST, streptactin-binding Strep-tag, streptavidin-binding peptide and other tags abound, and many incorporate magnetic bead technology.

On the protein fractionation and preparation front, in January Qiagen launched a new line of protein fractionation kits under the Qproteome brand. The range features kits, including reagents, buffers and columns, for such common tasks as phosphoprotein purification, glycoprotein fractionation and albumin depletion.

“The simplicity of the kits and procedures offers a gentle introduction to the world of protein science for molecular biologists who might be wary of having to learn new techniques or understand complex technologies,” says Cassing.

Many protein purification kits based on filtration are available in spin-column format, allowing faster processing. Vivascience, a subsidiary of Sartorius, based in Hanover, Germany, has developed a range of specialized spin-column purification kits, based on the firm’s membrane matrix of stabilized regenerated cellulose. Because of the porous structure, the surface available for contact and binding is about 100 times that of the same volume of traditional bead-based resins, allowing parallel separation

of proteins with high yields in less than 20 minutes.

The new ProteoSpin line of spin column kits for protein clean up from Norgen Biotek in St Catharines, Ontario, is based on a patent-pending technology using modified silicon carbide (SiC) as the matrix rather than the usual silica (SiO₂). SiC has all the benefits of silica resin and more, says Yousef Haj-Ahmad, president and chief executive of Norgen Biotek. The hydrophobic and hydrophilic surfaces of SiC can be exploited directly rather than having to chemically add active sites, as with a silica matrix. “The lack of porosity is an advantage because it enables the purification of a wide size range of proteins,” Haj-Ahmad notes. “With silica-based resins, even if they are ion-exchange type, one finds size restrictions for proteins because of the micropores.” Also, the unique way that SiC’s surface charges allow it to function as an ion exchanger means that salts are not needed to elute proteins in most cases. The first ProteoSpin kits are focused on protein preparation for downstream applications such as mass spectrometry.

The next challenge

Many of the large equipment providers are now concentrating on streamlining and speeding up the overall protein-processing workflow, from improving initial sample clarification through to a smoother transition to mass spectrometry, microarray technology or X-ray crystallography. Informatics at the back end also remains a challenge, in terms of analysing the results of large-scale analysis and feeding them back into the production process.

More sophisticated successors to the His₆ and GST tags commonly used in protein purification are also being sought. “Most people are still using the classic protein purification tags that have been around for many years,” says Roskey. “There probably need to be some new advances in that area, better ways to grab hold of proteins as you express them. That’s something that a lot of people are working on.”

But perhaps the biggest challenge is in applying the skills learned with soluble proteins to membrane proteins. “There’s not a soluble protein that we can’t purify for you,” says Simpson. “But once you get to membrane proteins, which are particularly interesting for the pharma industry, they’re a nightmare. They’re probably key receptors for most drugs but they’re designed to be insoluble in physiological structures. How to address membrane proteins is one of the critical bottlenecks. Everyone’s putting so much effort into it that it will be solved, and whoever solves it will really hit the pot of gold because it will change the face of our industry and the way medicines are designed.” ■

Tim Chapman is a freelance journalist based in Halifax, UK.



Large-scale protein production needs equally high-throughput analysis.

Company	Products/activity	Location	URL
Chromatography and affinity tags			
Abcam	Antibodies and affinity tags	Cambridge, UK	www.abcam.com
Active Motif	Antibodies; kits and reagents for cell fractionation; Ni-TED protein purification system for His ₆ -tagged proteins, nuclear protein purification	Carlsbad, California	www.activemotif.com
Alpha Diagnostics	Custom peptides and polyclonal antibodies; products for affinity chromatography	San Antonio, Texas	www.4adi.com
Biotech Support Group	Products for nucleic acid and protein isolation and removal; HPLC columns	East Brunswick, New Jersey	www.biotechsupportgroup.com
Covalys Biosciences	SNAP-Tag protein tags for protein fusions	Witterswil, Switzerland	www.covalys.com
IBA	Strep-tag/Strep-tactin affinity-purification system; One-STREP system for identifying protein complexes; protein expression and purification services	Göttingen, Germany	www.iba-go.com
MoBiTec	Products for molecular biology; products for affinity chromatography	Göttingen, Germany	www.mobitec.de
NanoPlex Technologies	Nanobarcodes, metallic nanoparticle-detection tags for proteomics and other applications	Mountain View, California	www.nanoplextech.com
Norgen Bioteck	ProteoSpin silicon carbide spin-column kits for protein purification	St Catharines, Ontario	www.norgenbiotek.com
PhyNexus	Single-use pipette tip chromatography columns for mini-scale protein purification	San Jose, California	www.phynexus.com
Pierce Biotechnology	Reagents and kits for protein chemistry, protein purification and chromatography	Rockford, Illinois	www.piercenet.com
Promega	Vectors, reagents and kits for genomics, proteomics and cellular analysis; magnetic resins for purification of His- and GST-tagged proteins	Madison, Wisconsin	www.promega.com
Qiagen	BioRobot multifunctional workstation; Two-Step affinity-purification system for expression, purification and detection of proteins; Qproteome kits for protein fractionation	Venlo, The Netherlands	www.qiagen.com
SEPIAtec	HPLC/SPE equipment for high-throughput purification and fractionation	Berlin, Germany	www.sepiatec.com
Sigma-Aldrich	Reagents and kits for life-sciences research; antibodies; chromatography media	St Louis, Missouri	www.sigmaaldrich.com
Stratagene	InterPlay purification systems for protein-protein interactions	La Jolla, California	www.stratagene.com
Tosoh Biosep	Resins and pre-packed columns for liquid chromatography	Stuttgart, Germany	www.tosohbiosep.com
Vivascience	Vivapure chromatography resins, spin columns and protein purification kits	Hannover, Germany	www.vivascience.com
Zymo Research	Reagents for molecular biology; His-Spin Protein Miniprep	Orange, California	www.zymoresearch.com
Electrophoresis equipment and gel-analysis software			
Advanced American Biotechnology & Imaging	Gel image analysis systems	Fullerton, California	www.aabi.com
Alpha Innotech	2D-gel analysis software; gel imaging and documentation systems	San Leandro, California	www.alphainnotech.com
Applied Maths	GelCompar and BioNumerics software for gel fingerprint analysis; bioinformatics services	Kortrijk, Belgium	www.applied-maths.com
BioSystematica	Gel-imaging software supplier	Tavistock, UK	www.biosystematica.com
Cambrex	Gels and products for molecular and cell biology	Walkersville, Maryland	www.cambrex.com
DECODON	Delta2D software for 2D-gel image analysis and information storage	Greifswald, Germany	www.decodon.de
Elchrom Scientific	Tools for protein research; gel electrophoresis, membrane for western blots, resins	Cham, Switzerland	www.elchrom.com
G Biosciences	FOCUS kits for proteome analysis; products for protein purification, 1D and 2D electrophoresis, assays, protein modification, cross linkers, dialysis, electro-elution	St Louis, Missouri	www.genotech.com
Genomic Solutions	Investigator 2D electrophoresis system; HT Analyzer and HT Analyzer Evolution 2D-gel analysis software	Ann Arbor, Michigan	www.genomicsolutions.com
Kapelan	LabImage 1D-gel analysis software	Halle/Saale, Germany	www.labimaging.com
Kodak	Imaging systems and image-analysis software; 1D Image Analysis Software	Rochester, New York	www.kodak.com/US/en/health/scientific/products/family.shtml
Media Cybernetics	Gel-Pro Analyzer gel-imaging software; image-analysis software	Silver Spring, Maryland	www.mediacy.com
Millipore	Automated equipment for molecular biology, biochemistry, genomics and proteomics; Montage in-gel digest kit; Amicon centrifuge filters	Bedford, Massachusetts	www.millipore.com
NextGen Sciences	A2DE automated 2D-gel electrophoresis workstation and A2DEoptimizer; expressionfactory automated gene-to-purified-protein expression system	Huntingdon, UK	www.nextgensciences.com
Nonlinear Dynamics	Phoretix image-analysis and database software for 1D and 2D gel electrophoresis and high-throughput arrays; Progenesis for the analysis of high-throughput 2D gels	Newcastle-upon-Tyne, UK	www.nonlinear.com
SciTech	Gel imaging and documentation systems	Preston, Victoria	www.scitech.com.au
Syngene	Dymension 2D-gel analysis software; gel imaging and documentation	Frederick, Maryland	www.syngene.com
UVP	1D-gel analysis system; gel imaging and documentation systems	Upland, California	www.uvp.com
Automated systems			
Agilent	2100 Bioanalyzer for genomics and proteomics; 5100 Automated Lab-on-a-Chip electrophoresis platform for high-throughput nucleic acid and protein analysis	Palo Alto, California	www.agilent.com
Applied Biosystems	Protein sequencers, mass spectrometers, chromatography systems, peptide synthesizers	Foster City, California	www.appliedbiosystems.com
BD Biosciences	FACS range of flow cytometers; antibody arrays; reagents for molecular biology	Franklin Lakes, New Jersey	www.bd.com
Beckman Coulter	Biomek plate-handling robot; automated tools, ProteomeLab PF2D	Fullerton, California	www.beckmancoulter.com
Bio-Rad	Bio-Plex system for multiplexing antibody-type assays; protein purification products; Experion automated electrophoretic separation system for proteins and nucleic acids	Hercules, California	www.bio-rad.com
Caliper Life Sciences	LabChip microfluidic systems for protein and DNA separation; LabChip 90 for high-throughput protein assay	Hopkinton, Massachusetts	www.calipertech.com
Fluidigm	Microfluidics chip for protein crystallization	South San Francisco, California	www.fluidigm.com
GE Healthcare	Instruments, equipment and products for genomics, proteomics and cellular assays; ÄKTA Express high-throughput automated liquid chromatography platform; ImageMaster 2D-gel analysis software	Little Chalfont, UK	www.amershambiosciences.com
Gilson	Automated liquid-handling, pipetting and protein crystallization instruments	Middleton, Wisconsin	www.gilson.com
Gyros	Cyberlab automated workstation for protein crystallization		
	CD microlaboratory for sample preparation for protein mass spectrometry	Uppsala, Sweden	www.gyros.com

Company	Products/activity	Location	URL
Hamilton Company	MicroLab STAR automated liquid-handling workstations for proteomics applications	Reno, Nevada	hamiltoncomp.com
Hitachi High-Technologies	Automated instrumentation for high-throughput protein purification and analysis	Tokyo, Japan	www.hitachi-hitec.com
Magnetic Biosolutions	Robotic workstations for automated biomagnetic separation	Stockholm, Sweden	www.magbio.com
PerkinElmer Life Sciences	HydroGel BioChip for DNA and protein arrays; MultiPROBE II liquid-handling workstation, accessories and kits for protein purification; ProXCISION gel-cutting robot	Boston, Massachusetts	las.perkinelmer.com
Proteodyne	BioCube automated systems	Windsor, Connecticut	www.proteodyne.com
Tecan	Genesis and Freedom EVO liquid-handling workstations	Männedorf, Switzerland	www.tecan.com
Thermo Electron Corporation	KingFisher system for magnetic-particle-based protein and nucleic acid purification	Waltham, Massachusetts	www.thermo.com
Whatman	Separation technology, microplates, Biometra instruments and components for automated systems	Maidstone, UK	www.whatman.com

Magnetic bead technology

Agarose Bead Technologies	Agarose beads, magnetic beads, metal-chelating beads	Tampa, Florida	www.abtbeads.com
Bio-Nobile	PickPen hand-held magnetic bead separation equipment and kits for nucleic acid isolation and protein purification	Turku, Finland	www.bio-nobile.com
BioScience Beads	Agarose beads, magnetic beads, hydrophobic beads for hydrophobic affinity chromatography	West Warwick, Rhode Island	www.bioscience-beads.com
chemagen	PVA-based superparamagnetic beads	Baesweiler, Germany	www.chemgen.de
Dynal Biotech	Dynabeads superparamagnetic beads for immunoprecipitation and phage-display panning, protein isolation and purification	Oslo, Norway	www.dynalbiotech.com
New England Biolabs	Reagents and kits for molecular biology; streptavidin magnetic beads	Beverly, Massachusetts	www.neb.com
Novagen	Reagents and kits for molecular biology; Robopop protein purification kits; Magnetight separation stand for 96-well plate magnetic bead systems	Madison, Wisconsin	www.novagen.com
Polysciences	BioMag superparamagnetic beads; Amberlite polymers	Warrington, Pennsylvania	www.polysciences.com

Services

Advion BioSciences	Contract liquid chromatography and services for mass spectrometry	Ithaca, New York	www.advion.com
Ambit Biosciences	ProteomeScan platform for drug target discovery and validation via phage display and affinity chromatography	San Diego, California	www.ambitbio.com
Australian Proteome Analysis Facility (APAF)	Contract research and product development for proteomics research	Sydney, New South Wales	www.proteome.org.au
BioGenes	Customized monoclonal and polyclonal antibodies; immunoassays; peptides	Berlin, Germany	www.biogenes.de
Lonza Group	Contract manufacture of monoclonal antibodies and recombinant proteins from mammalian cell culture	Basel, Switzerland	www.lonza.com
Paragon Bioservices	Recombinant protein expression, protein purification	Baltimore, Maryland	www.paragonbioservices.com
Protagen	Protein analysis services	Dortmund, Germany	www.protagen.com
ProteinX	Protein purification and analysis; biomolecular interaction analysis using Biacore 1000	San Diego, California	proteinx.com
Proteome Sciences	ProteoSHOP system for protein purification and mass spectrometry; Sensitizer reagents for mass spectrometry; disease biomarker discovery	Cobham, UK	www.proteomics.com
psf biotech	Services for drug discovery through structural genomics	Berlin, Germany	www.psf-ag.com
Trenzyme	Genomics and proteomics services: cloning, mutagenesis, library construction, gene expression, protein purification	Konstanz, Germany	www.trenzyme.com
Upstate Group	Bulk protein production and purification	Charlottesville, Virginia	www.upstate.com

General

Affibody	High-throughput protein identification and characterization by proprietary binding proteins	Stockholm, Sweden	www.affibody.com
Apogent Technologies	Labware and equipment for the life sciences	Portsmouth, New Hampshire	www.apogent.com
Asys Hitech	Microplate readers, washers and dispensers: Flexispense HT 8 channel and 16-channel dispensers for 384- and 1,536-well plates	Eugendorf, Austria	www.asyshitech.com
BMG LABTECH	Microplate and array readers and handling systems	Offenburg, Germany	www.bmg-labtechnologies.com
Brinkmann	Laboratory instrument suppliers; software; consumables	Westbury, New York	www.brinkmann.com
Calbiochem	ProteoExtract protein extraction kits, including subcellular proteome	San Diego, California	www.calbiochem.com
Chemicon	Reagents and kits for antibody labelling and protein purification and detection	Temecula, California	www.chemicon.com
EPICENTRE	Products for genomics, proteomics and protein purification	Madison, Wisconsin	www.epibio.com
Eppendorf	Laboratory instrumentation and consumables for molecular and cell biology; Eppendorf PerfectPure C-18 Tips for peptide purification for MS analysis	Hamburg, Germany	www.eppendorf.com
GenScript	Supplier of kits and services for cell biology and molecular biology	Piscataway, New Jersey	www.genscript.com
Invitrogen	Kits and reagents for cloning, genomics, proteomics, molecular and cell biology; antibodies, fluorescent-tagged antibodies; human proteome microarray	Carlsbad, California	www.invitrogen.com
LCI	STS Biochip platform for sample preparation for MALDI-MS	Fremont, California	www.lumicyte.com
Nanodrop	Small sample spectrophotometry	Wilmington, Delaware	www.nanodrop.com
New England Biolabs	Reagents and kits for molecular biology; streptavidin magnetic beads	Beverly, Massachusetts	www.neb.com
Qbiogene	Products, kits and reagents for molecular and cell biology	Irvine, California	www.qbiogene.com
R&D Systems	Reagents and kits for molecular and cell biology	Minneapolis, Minnesota	www.RnDSystems.com
Roche Diagnostics	Reagents and kits for molecular biology, functional genomics and proteomics research; Rapid Translation System scalable <i>in vitro</i> transcription/translation protein-expression	Lewes, UK	www.roche-applied-science.com
TaKaRa Bio	Reagents, kits and services for molecular biology, genomics and proteomics research	Shiga, Japan	www.takara-bio.co.jp/english
Wako Chemicals	Speciality chemicals, bioproducts and clinical diagnostic reagents	Richmond, Virginia	www.wakousa.com

● See advertisement

Contacts

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

US Head Office, New York

345 Park Avenue South, 10th Floor,
New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager/ Corporations:

Peter Bless

Classified Sales Representatives:

Phone: +1 800 989 7718

New York/ Pennsylvania/

Latin America:

Kelly Roman

Midwest USA/ Maryland/ NIH:

Wade Tucker

East USA/ Canada:

Janine Taormina

San Francisco Office

Classified Sales Representative:

Michaela Bjorkman

West USA/ West Corp. Canada

225 Bush Street, Suite 1453

San Francisco, CA 94104

Tel +1 415 781 3803

Fax +1 415 781 3805

e-mail: m.bjorkman@nature.sf

European Head Office, London

The Macmillan Building, 4 Crinan Street,
London N1 9XW, UK

Tel +44 (0) 20 7843 4961

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs Sales Director:

Nevin Bayoumi (4978)

European Sales Manager:

Andy Douglas (4975)

Advertising Production

Manager: Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Niamh Shields

European Satellite Office

Germany/ Austria/ Italy/

The Netherlands/ Belgium:

Patrick Phelan

e-mail: p.phelan@nature.com

Reya Silao

e-mail: rsilao@nature.com

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Harai Katamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami

e-mail: rasami@nature.jp

naturejobs

Guiding hands

Imagine what would happen if scientists' professional advancement was based not on the number of papers they had published but on how many people's careers they had helped. This might actually be a better measure of success than authorship alone, because good mentoring means that a scientist could indirectly have a hand in a paper, or at least influence an investigator's approach, a generation after he or she had retired.

But back to reality. Demand for quality guidance is huge. A forum facilitated by the US National Academy of Sciences suggested that good mentoring — not stipends — was the most important factor influencing the careers of US postdocs. And last month, meetings by the European Council of Doctoral Candidates and Junior Researchers and the US National Postdoctoral Association both highlighted the need for more guidance.

To encourage more and better guidance, *Nature* and the UK National Endowment for Science, Technology and the Arts this year launched an award scheme for scientific mentoring. The first two winners, Tom Kibble of Imperial College London and Innes Cuthill of the University of Bristol (see page 802), have set an example that other scientists would do well to follow. And, by doing so, the pair have advanced their own careers. By offering fair guidance rather than fostering ruthless competition, Kibble and Cuthill have maintained working relationships with their students long after their protégés have left their labs.

Kibble and Cuthill may be exceptional, but they are doubtless not alone. Other great mentors are out there, and many institutions have schemes to laud them. But few outside the university hear about these awards, and few administrators use them as the basis for tenure or promotion. In the coming months, *Naturejobs* will promote other mentoring awards, so e-mail naturejobseditor@naturedc.com to suggest any such awards that we should highlight.

Paul Smaglik
Naturejobs editor



Contents

SPECIAL REPORT

Model mentors p802

CAREER VIEW

Nuts & Bolts
International ventures
Graduate Journal
Goodbye to romance
Movers
Ram Sasisekharan p804

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

We all know a special person who has inspired us. Paul Smaglik learns from the winners of the first UK award for excellent mentoring in science.

What makes a good supervisor? Someone who allows students to ask questions, no matter how silly they might seem? Someone who lets students grapple with problems before jumping in with answers and explanations? Or someone who emphasizes the positive, even when an experiment goes awry?

The winners of the first *Nature*/UK National Endowment for Science, Technology and the Arts (NESTA) awards for mentoring in science have all these qualities. The winner of the lifetime mentoring award is Tom Kibble, senior research fellow and emeritus professor of theoretical physics at Imperial College, London. The mid-career mentoring award was presented to Innes Cuthill, professor of behavioural ecology at the University of Bristol. The winners demonstrate that intellect and humanity can combine to advance both the mentor's and the protégé's careers, say the scientists who proposed them for this award.

Kibble's nominators describe him as modest, brilliant and always approachable. Even when Kibble's administrative duties and importance grew during his career, he still found time for his students. "This approachability was very important; it is easy to feel intimidated by a person of such stature," says Danièle Steer, a lecturer at the University of Paris VII.

She recalls a conversation she had with Kibble soon after beginning her thesis work. "I had that sinking feeling that I did not understand anything and would never be able to understand anything," she says. Kibble sensed that something was amiss. He told her there were many things he didn't understand either. "He was completely sincere," says Steer. "That small comment helped me during my thesis, and indeed ever since."

Ed Copeland, professor of theoretical physics at the University of Sussex, says Kibble would always make time for students — even if they weren't his. Towards the end of his PhD, Copeland was interviewed for a position at Imperial College. After his seminar was over, Copeland asked to meet Kibble. "Although clearly

very busy, and not really working on the topic of my research, he spoke with me for half an hour, asking me about my work, telling me about his, and finally offering some advice about delivering seminars."

Besides advising the young scientist not to use yellow pens during seminars, Kibble helped him learn when a problem is worth solving, realize when it could lead to a brick wall, and when to look for alternative routes around it.

Mark Hindmarsh, reader in theoretical physics

at the University of Sussex, also benefited from Kibble's instincts and astute diplomacy. When Hindmarsh came to Imperial College as a PhD student, he was determined to study quantum gravity — even though he didn't really understand what it was. "I am deeply grateful to him for gently steering me away from it (without ever issuing instructions) into the far greener pastures of particle cosmology," Hindmarsh says. It was a wise decision; his first-choice area has seen little progress, whereas the field that Kibble directed him into has flourished.

Perhaps Kibble's greatest achievements as a mentor are the workshops he organized that taught

science and helped younger scientists to serve as mentors to each other. During his early years at Imperial College, he initiated a weekly discussion group on particle cosmology and related fields. It grew from a few dozen people from the college to drawing over 60 from around Britain, says Neil Turok, chair of mathematical physics at the University of Cambridge. Kibble's open-door policy, as well as the promotion of his discussion group, benefited both sides: young scientists gained papers early in their careers, and Kibble's own publication record took off.

Kibble's achievements stem from his ability to foster individual relationships, a skill he shares with his fellow winner. Cuthill, say his nominators, has an uncanny ability to treat scientists as individuals, and to tailor training to different needs and personalities.

John Swaddle, associate professor at the College of William and Mary in Williamsburg, Virginia, recalls



Words of encouragement: Tom Kibble, winner of the lifetime mentoring award, always found time to help his students.



Nancy Rothwell presents mid-career award winner Innes Cuthill with his prize (left) during the celebratory party at the Science Museum, London (below).



H. MENTZ being taking aback in graduate school when Cuthill left him alone but gave another graduate student specific reading assignments and a detailed set of research questions to explore. "At the time I thought I had it easy, as Innes was not asking me to do these assignments, and that the other student was being treated more harshly," Swaddle says. "With a couple of years' hindsight, I realized that Innes was merely motivating this other student in a more appropriate way for their adviser-student relationship." And he realized that Cuthill was hands-off with him because he was heading in the right direction.

Sarah Hunt, Young People's Programme Development Officer at the British Association for the Advancement of Science, says that Cuthill is so good at understanding each scientist's "idiosyncrasies" that he can understand what motivates different individuals and adjust his approach to each person as they develop. "I have watched other colleagues being moulded to specific ways of working and thinking," she says. But that wasn't Cuthill's way of doing things.

Hunt recalls her first day of graduate school. Cuthill showed her to her desk, then said he didn't care what time she turned up in the morning, as long as she got her work done in the end. "As it happens, I've never worked harder, and never enjoyed it so much," she says.

During Hunt's first year, Cuthill played a big role in her development. He helped her choose experiments, taught her how to work with animals and helped her learn to use statistics. As a result, she had two papers in press by the end of her first year. "Innes then took a back seat and allowed me to be independent," Hunt says.

"I am deeply grateful to Kibble for gently steering me away from studying quantum gravity into the far greener pastures of particle cosmology"

— Mark Hindmarsh

Cuthill's patience and non-judgemental attitude also proved helpful for the many other young scientists he has mentored. Sasha Dall, lecturer in mathematical ecology at the University of Exeter in Cornwall, calls him the "single most influential person in my professional life" — and not just for the entertaining lectures that Cuthill delivered, which remind him that teaching is an important part of being a researcher. "He had (and still has) the knack of listening to my often incoherent ramblings and gently revealing the useful ideas buried deep within via very constructive criticism," Dall says.

Jonathan Wright, professor of biology at the Norwegian University of Science and Technology, Trondheim, says that Cuthill doesn't have any single formula for successful mentoring. "He just seems to quietly get it right on a one-to-one basis with each of his students," he explains.

That sounds like a good approach for other mentors to follow.

Paul Smaglik is editor of *Naturejobs*.

NUTS & BOLTS

GRADUATE JOURNAL

Goodbye to romance

During my medical studies I loved reading anecdotes about the life and work of eminent scientists. How they came across their famous discoveries in the cloistered atmosphere of distinguished universities. How intimate scientific circles seemed to pose an enigmatic problem, contemplate it — sometimes for years — and then solve it elegantly. How these idols wove science, life and companionship together.

Today I'm disillusioned about the conditions under which research is carried out. Science seems to be developing into a business, interested in a quick and focused 'return on investment'. Within the short time of their contracts, researchers have to come up with results at all costs, including exhaustion or loss of privacy. Competition with other groups and dependence on external funding generates a huge pressure that is far from the scientific arcadia of my fantasies. Actual science doesn't resemble what can be read about it in biographies.

My graduate programme aims for the best it can do. It furnishes me with the skills that this science business requires: broad knowledge of the field, scientific writing, an array of methodologies, diligence and persistence. It can't change the face of research. But the beauty of science itself is left untouched. I hope to retain at least some of this naivety. ■

Tobias Langenhan is a first-year graduate student in neuroscience at the University of Oxford, UK.

International ventures

Finding a job abroad adds distance to the string of challenges job-seekers face. How do you get to the people you need to reach? Will you be considered if you apply from another city, let alone another continent? Should you move to the location first and then launch your search?

Start at home by researching the place you want to work, getting a clear idea of employers you are interested in and building relationships. Combine information-gathering with online activities as well as engaging your university librarian. These activities will provide the foundation for your search.

Next, begin to craft your correspondence. In a long-distance search, written and telephone communications carry added weight. Don't cut corners on your marketing package. Make sure that your resumé or



**With Deb Koen
Careers consultant**

CV is professionally presented (see *Nature* 427, 570; 2004), your e-mails and letters are error-free, and your telephone skills are polished (see *Nature* 429, 584; 2004).

Next, network at home. You might secure a position with a local organization that has operations in your targeted area, or join an international concern that has a satellite office nearby. Proving yourself in your current location might secure an assignment at a branch in another country. At the very least, you can get in touch with colleagues abroad. And don't forget to

seek referral from advisers or supervisors with ties to where you want to go.

These preliminary steps create the foundation for a strategic visit. E-mail or ring organizations to let them know you will be there soon. Your contacts will take you seriously if it's clear that your move is imminent. Ideally, plan a two-week visit with some interviews lined up and spend the rest of the time setting up further meetings.

Once there, make the process as seamless as possible: know any legal requirements and show potential employers your familiarity with the culture and the language.

An international job search takes determination and time, but combining a long-distance campaign with personal follow-up will move you closer to your destination. ■

Deb Koen is vice-president of Career Development Services and a columnist for *The Wall Street Journal's CareerJournal.com*.

MOVERS

Ram Sasisekharan, chief technology adviser, MVM Life Science, Cambridge, Massachusetts



Not many people would sink their teeth into sequencing sugar structures when the human genome held such mouth-watering promise. But Ram Sasisekharan sought and accepted a challenge from his thesis adviser, Robert Langer at Massachusetts Institute of Technology (MIT).

Sasisekharan found the long-ignored sugars field — severely hampered by a lack of tools and technology — an opportunity in disguise. Unlike DNA,

sugars can't be amplified through tools such as the polymerase chain reaction. They are also made up of complex structures in varied abundance. He used a number-based approach to identify the different pieces, allowing mathematical manipulations to solve the structure. "Sequencing DNA is like walking through a ladder, but sequencing sugars is more like putting a puzzle together," he says.

The puzzle complete, he and Langer set up Momena Pharmaceuticals. With a detailed structural knowledge of polysaccharides, it is possible to improve existing drugs, create generics and discover novel drugs by better understanding their role in cell function.

With \$83 million raised, Momena is launching a heparin drug and delving into cancer. By discovering how cancer cells gain a growth advantage by altering their sugar coat, Sasisekharan hopes to develop insights for a potential new avenue of therapeutics.

Stressing the importance of mentors, Sasisekharan notes that meeting Langer was a career-changing event. Although a graduate student at Harvard University, he heard Langer give a talk during a chance visit to MIT. Langer's 'can-do' approach to science inspired him to be passionate about big-picture problems.

"The most exciting thing is to be able to bring a team effort to a complex problem," he says. His latest team is MVM Life Science Partners, the US subsidiary of the UK-based life-sciences investment firm.

The move towards an integrated, systems-biology approach to science is a welcome change for Sasisekharan. He encourages young scientists to gain exposure to challenging problems that are also practical in nature.

In doing so, he's turned seemingly useless molecules into potential therapeutic drugs. It can't get much sweeter than that. ■

CV **2003–05:** Professor, biological-engineering division, Massachusetts Institute of Technology, Cambridge, Massachusetts.

2003–current: Member, joint steering committee, Momena Pharmaceuticals/Sandoz-Novartis Venture, Cambridge, Massachusetts.

2001–current: Co-founder and board member, Momena Pharmaceuticals, Cambridge, Massachusetts.

1996–current: Core member, NanoTechnology Lab, Massachusetts Institute of Technology.

Ivory tower

A place to call our own.

Bruce Sterling

Our problem was simple. We needed an academy, but professional careers in conventional science were out of the question for us. We were 10,000 physicists, entirely self-educated on the Internet.

Frankly, physics is a lot easier to learn than physicists used to let on. The ultimate size of the smallest particles, the origin and fate of the Universe — come on, who could fail to take a burning interest in those subjects? If we were genuinely civilized, that's all we would talk about. In the new world of open access, ultrawide broadband and gigantic storage banks, physics is just sort of sitting there. It's like a vast intellectual Tinkertoy! We cranky net-geeks had to find a way to devote every waking moment to our overpowering lust for physics. Of course, we demanded state support for our research efforts (just like real scientists do) but, alas, the bureaucrats wouldn't give us the time of day.

So to find time for our kind of science, we had to dump a few shibboleths. For instance, we never bother to 'publish': we just post our findings on weblogs, and if they get a lot of links, hey, we're the Most Frequently Cited. Tenure? Who needs that? Never heard of it! Doctorates, degrees, defending a thesis — don't know, don't need 'em, can't even be bothered!

Organizing ourselves was a snap. If you are a maths genius whose primary language is Malayalam and whose main enthusiasm is wave-particle duality, you stand out on the net like a buzzing hornet in a spiderweb. You're one in a million, pal — but in a world of ten billion people, there's 10,000 of us. We immediately started swapping everything we knew on collaborative weblogs.

As most of us were Indian and/or Chinese (most of everybody is Indian and/or Chinese) we established our Autodidacts' Academy on the sun-baked, sandstone flats of the desert of Rajasthan, not too far from the deserted Mughal utopia of Fatehpur Sikri. We were dreamy, workaholic utopians trying to wrest a living out of barren wilderness. Something like Mormons, basically. However, as it was the 2050s, we also had unlimited processing power, bandwidth, search engines, social software and open-source everything. How could we fail?

Basically, we recast human existence as a bioengineering problem. How do you move



enough nutrient through human brain tissue to allow an entire city of people to blissfully contemplate supersymmetric M-branes? The solutions were already scattered through the online technical literature; we just Googled it all up and set it to work. Our energy is solar; water is distilled and recycled; and the ivory gleaming domes and spires of our physics ashram are computer-fabricated grit, glue and sawdust. All our lab equipment is made of garbage.

Our visitors are astounded to see (for instance) repurposed robotic vacuum cleaners equipped with tiller blades digging out our 150-kilometre accelerator tunnel. But why not? In the 2050s, even the junk is ultra-advanced, and nobody knows how to repair it. Any sufficiently advanced garbage is indistinguishable from magic.

Our daily diet, which is free of charge, is fully defined Physicist Chow. It's basically sewage, with its bioenergetic potential restored by genetically altered yeasts. Some diners fail to appreciate the elegant mathematical simplicity of this solution to the age-old problem of a free lunch. But if they don't get it, then they don't belong here with us, anyway.

There's no money and no banking here. Instead, every object is tracked by RFID tags and subjected to a bioenergetic, cost-benefit, eBay-style arbitrage by repurposed stock-market buy-sell software agents. In practice, this means that when you need something

new, you just pile up the things you don't want by your doorway until somebody shows up and gives you the thing you do want. Economists who visit here just flee screaming — but come on, was economics ever really a 'science'? We're with Rutherford: it's physics or it's stamp collecting!

You might imagine that women would find our monastic, geeky life unattractive, but our academy's crawling with co-eds. A few are female physicists — the usual proportion — but the rest are poets, lit. majors, anthropologists and gender studies mavens. These gals showed up to condemn our reductionist, instrumental male values, but they swiftly found out that our home is ideal for consciousness-raising, encounter groups and performance art. So women now outnumber us three to two. That's not a problem. We don't bother them with our weird obsessions, they don't bother us with theirs, and whatever happens between us after dark is nobody's business.

We have a beautiful, spiritual thing going on here. Feel free to join us. Please, no more atomic-bomb fans. We know that atomic bombs are a dead simple, 100-year-old technology. Anybody with a search engine, half a brain and a lot of time can tinker one up. But really, why even bother? It's beneath us! ■

Bruce Sterling is currently 'visionary in residence' at the Art Center College of Design in Pasadena. He also writes science fiction novels and maintains a daily weblog at <http://blog.wired.com/sterling>.